FAMOUS PROBLEMS
AND OTHER MONOGRAPHS

FAMOUS PROBLEMS OF ELEMENTARY GEOMETRY
BY F. KLEIN

FROM DETERMINANT TO TENSOR
BY W. F. SHEPPARD

INTRODUCTION TO COMBINATORY ANALYSIS
BY P. A. MACMAHON

THREE LECTURES ON FERMAT'S LAST THEOREM
BY L. J. MORDELL

AMS CHELSEA PUBLISHING
American Mathematical Society • Providence, Rhode Island

FAMOUS PROBLEMS OF ELEMENTARY GEOMETRY
was originally published by Ginn and Company

FROM DETERMINANT TO TENSOR
was originally published by Oxford University Press

INTRODUCTION TO COMBINATORY ANALYSIS
was originally published by Cambridge University Press

THREE LECTURES ON FERMAT'S LAST THEOREM
was originally published by Cambridge University Press

2000 *Mathematics Subject Classification.* Primary 00B10.

Library of Congress Catalog Card Number 62-18132
International Standard Book Number 0-8218-2674-3

FIRST EDITION 1955
SECOND EDITION 1962, 1980

10 9 8 7 6 5 4 3 2 1 04 03 02 01 00

EDITOR'S PREFACE

This work, like its companion volume, *Squaring the Circle, and other Monographs* by Hobson et al., consists of a reprint in one volume of several books on mathematics that were originally published as separate volumes.

The reason for the selection of the four books that comprise this volume is that each is a valuable and important work and that each is of interest to a fairly wide circle of mathematicians and students.

The reason for their inclusion in a single volume is neither learned nor recondite. The reason is purely economic: Reprinted separately, the books would have to be priced at not much less than the price of the whole present volume (if they could be so reprinted at all). Anyone who buys the book for the sake of one of the four volumes that it contains will surely find the other three of interest and will consider them to be a worthwhile and welcome addition to his library.

FAMOUS PROBLEMS

OF

ELEMENTARY GEOMETRY

THE DUPLICATION OF THE CUBE
THE TRISECTION OF AN ANGLE
THE QUADRATURE OF THE CIRCLE

AN AUTHORIZED TRANSLATION OF F. KLEIN'S
VORTRÄGE ÜBER AUSGEWÄHLTE FRAGEN DER ELEMENTARGEOMETRIE
AUSGEARBEITET VON F. TÄGERT

BY

WOOSTER WOODRUFF BEMAN
1850—1922

AND

DAVID EUGENE SMITH
EMERITUS PROFESSOR OF MATHEMATICS IN COLUMBIA UNIVERSITY

SECOND EDITION REVISED, AND ENLARGED WITH NOTES

BY

RAYMOND CLARE ARCHIBALD
PROFESSOR OF MATHEMATICS IN BROWN UNIVERSITY

PREFACE.

THE more precise definitions and more rigorous methods of demonstration developed by modern mathematics are looked upon by the mass of gymnasium professors as abstruse and excessively abstract, and accordingly as of importance only for the small circle of specialists. With a view to counteracting this tendency it gave me pleasure to set forth last summer in a brief course of lectures before a larger audience than usual what modern science has to say regarding the possibility of elementary geometric constructions. Some time before, I had had occasion to present a sketch of these lectures in an Easter vacation course at Göttingen. The audience seemed to take great interest in them, and this impression has been confirmed by the experience of the summer semester. I venture therefore to present a short exposition of my lectures to the Association for the Advancement of the Teaching of Mathematics and the Natural Sciences, for the meeting to be held at Göttingen. This exposition has been prepared by Oberlehrer Tägert, of Ems, who attended the vacation course just mentioned. He also had at his disposal the lecture notes written out under my supervision by several of my summer semester students. I hope that this unpretending little book may contribute to promote the useful work of the association.

F. KLEIN.

GÖTTINGEN, Easter, 1895.

TRANSLATORS' PREFACE.

At the Göttingen meeting of the German Association for the Advancement of the Teaching of Mathematics and the Natural Sciences, Professor Felix Klein presented a discussion of the three famous geometric problems of antiquity, — the duplication of the cube, the trisection of an angle, and the quadrature of the circle, as viewed in the light of modern research.

This was done with the avowed purpose of bringing the study of mathematics in the university into closer touch with the work of the gymnasium. That Professor Klein is likely to succeed in this effort is shown by the favorable reception accorded his lectures by the association, the uniform commendation of the educational journals, and the fact that translations into French and Italian have already appeared.

The treatment of the subject is elementary, not even a knowledge of the differential and integral calculus being required. Among the questions answered are such as these: Under what circumstances is a geometric construction possible? By what means can it be effected? What are transcendental numbers? How can we prove that e and π are transcendental?

With the belief that an English presentation of so important a work would appeal to many unable to read the original,

Professor Klein's consent to a translation was sought and readily secured.

In its preparation the authors have also made free use of the French translation by Professor J. Griess, of Algiers, following its modifications where it seemed advisable.

They desire further to thank Professor Ziwet for assistance in improving the translation and in reading the proof-sheets.

W. W. BEMAN.
D. E. SMITH.

August, 1897.

EDITOR'S PREFACE.

Within three years of its publication thirty-five years ago Klein's little work was translated into English, French, Italian, and Russian[1]. In the United States it filled a decided need for many years, and not a few teachers regretted that the work was allowed to go out of print. No other work supplied in such compact form just the information here found. Hence it seemed desirable to have a new edition with at least some of the slips of the first edition rectified, and with added notes illuminating the text.

The corrections and notes of the present edition are little more than revised extracts from my article in *The American Mathematical Monthly*[2], 1914. I am indebted to the Editors for courteously allowing the reproduction of this material.

R. C. A.

February, 1930.

[1] French translation by Griess, Paris, Nony, 1896; Italian by Giudice, Turin, Rosenberg e Sallier, 1896; Russian by Parfentiev and Sintsov, Kazan, 1898. This last translation seems to have been unknown to the editors of Klein's *Abhandlungen* (see v. 3, 1923, p. *28*).

[2] Remarks on Klein's "Famous Problems of Elementary Geometry", v. 21, p. 247—259.

CONTENTS.

INTRODUCTION.

PART I.

The Possibility of the Construction of Algebraic Expressions.

CHAPTER IV. THE CONSTRUCTION OF THE REGULAR POLYGON OF 17 SIDES.

CHAPTER V. GENERAL CONSIDERATIONS ON ALGEBRAIC CONSTRUCTIONS.

PART II.

Transcendental Numbers and the Quadrature of the Circle.

CHAPTER I. CANTOR'S DEMONSTRATION OF THE EXISTENCE OF TRANSCENDENTAL NUMBERS.

CHAPTER II. HISTORICAL SURVEY OF THE ATTEMPTS AT THE COMPUTATION AND CONSTRUCTION OF π.

CHAPTER III. THE TRANSCENDENCE OF THE NUMBER e.

INTRODUCTION.

This course of lectures is due to the desire on my part to bring the study of mathematics in the university into closer touch with the needs of the secondary schools. Still it is not intended for beginners, since the matters under discussion are treated from a higher standpoint than that of the schools. On the other hand, it presupposes but little preliminary work, only the elements of analysis being required, as, for example, in the development of the exponential function into a series.

We propose to treat of geometrical constructions, and our object will not be so much to find the solution suited to each case as to determine the *possibility* or *impossibility* of a solution.

Three problems, the object of much research in ancient times, will prove to be of special interest. They are

1. *The problem of the duplication of the cube* (also called the *Delian problem*).
2. *The trisection of an arbitrary angle.*
3. *The quadrature of the circle, i.e.*, the construction of π.

In all these problems the ancients sought in vain for a solution with straight edge and compasses, and the celebrity of these problems is due chiefly to the fact that their solution seemed to demand the use of appliances of a higher order. In fact, we propose to show that a solution by the use of straight edge and compasses is impossible.

The impossibility of the solution of the third problem was demonstrated only very recently. That of the first and second is implicitly involved in the Galois theory as presented to-day in treatises on higher algebra. On the other hand, we find no explicit demonstration in elementary form unless it be in Petersen's text-books, works which are also noteworthy in other respects.

At the outset we must insist upon the difference between *practical* and *theoretical* constructions. For example, if we need a divided circle as a measuring instrument, we construct it simply on trial. Theoretically, in earlier times, it was possible (*i.e.*, by the use of straight edge and compasses) only to divide the circle into a number of parts represented by 2^n, 3, and 5, and their products. Gauss added other cases by showing the possibility of the division into parts where p is a prime number of the form $p = 2^{2^\mu} + 1$, and the impossibility for all other numbers. No practical advantage is derived from these results; ***the significance of Gauss's developments is purely theoretical.*** The same is true of all the discussions of the present course.

Our fundamental problem may be stated: ***What geometrical constructions are, and what are not, theoretically possible?*** To define sharply the meaning of the word "construction," we must designate the instruments which we propose to use in each case. We shall consider

1. Straight edge and compasses,
2. Compasses alone,
3. Straight edge alone,
4. Other instruments used in connection with straight edge and compasses.

The singular thing is that elementary geometry furnishes no answer to the question. We must fall back upon algebra and the higher analysis. The question then arises: How

shall we use the language of these sciences to express the employment of straight edge and compasses? This new method of attack is rendered necessary because elementary geometry possesses no general method, no *algorithm*, as do the last two sciences.

In analysis we have first *rational* operations: addition, subtraction, multiplication, and division. These operations can be directly effected geometrically upon two given segments by the aid of proportions, if, in the case of multiplication and division, we introduce an auxiliary unit-segment.

Further, there are *irrational* operations, subdivided into *algebraic* and *transcendental*. The simplest algebraic operations are the extraction of square and higher roots, and the solution of algebraic equations not solvable by radicals, such as those of the fifth and higher degrees. As we know how to construct $\sqrt{ab}$, rational operations in general, and irrational operations involving only square roots, can be constructed. On the other hand, every *individual* geometrical construction which can be reduced to the intersection of two straight lines, a straight line and a circle, or two circles, is equivalent to a rational operation or the extraction of a square root. In the higher irrational operations the construction is therefore impossible, *unless we can find a way of effecting it by the aid of square roots.* In all these constructions it is obvious that the number of operations must be limited.

We may therefore state the following fundamental theorem: *The necessary and sufficient condition that an analytic expression can be constructed with straight edge and compasses is that it can be derived from the known quantities by a finite number of rational operations and square roots.*

Accordingly, if we wish to show that a quantity cannot be constructed with straight edge and compasses, we must prove that the corresponding equation is not solvable by a finite number of square roots.

A fortiori the solution is impossible when the problem has *no* corresponding algebraic equation. An expression which satisfies no algebraic equation is called a transcendental number. This case occurs, as we shall show, with the number π.

PART I.

THE POSSIBILITY OF THE CONSTRUCTION OF ALGEBRAIC EXPRESSIONS.

CHAPTER I.

Algebraic Equations Solvable by Square Roots.

The following propositions taken from the theory of algebraic equations are probably known to the reader, yet to secure greater clearness of view we shall give brief demonstrations.

If x, *the quantity to be constructed, depends only upon rational expressions and square roots, it is a root of an irreducible equation* $\phi(x) = 0$, *whose degree is always a power of* 2.

1. To get a clear idea of the structure of the quantity x, suppose it, *e.g.*, of the form

$$x = \frac{\sqrt{a + \sqrt{c + ef}} + \sqrt{d + \sqrt{b}}}{\sqrt{a} + \sqrt{b}} + \frac{p + \sqrt{q}}{\sqrt{r}},$$

where a, b, c, d, e, f, p, q, r are rational expressions.

2. The number of radicals one over another occurring in any term of x is called the *order of the term*; the preceding expression contains terms of orders 0, 1, 2.

3. Let μ designate the *maximum order*, so that no term can have more than μ radicals one over another.

4. In the example $x = \sqrt{2} + \sqrt{3} + \sqrt{6}$, we have three expressions of the first order, but as it may be written

$$x = \sqrt{2} + \sqrt{3} + \sqrt{2} \cdot \sqrt{3},$$

it really depends on only two distinct expressions.

We shall suppose that this reduction has been made in all the terms of x, *so that among the* n *terms of order* μ *none can be expressed rationally as a function of any other terms of order* μ *or of lower order.*

We shall make the same supposition regarding terms of the order $\mu - 1$ or of lower order, whether these occur explicitly or implicitly. This hypothesis is obviously a very natural one and of great importance in later discussions.

5. Normal Form of x.

If the expression x is a sum of terms with different denominators we may reduce them to the same denominator and thus obtain x as the quotient of two integral functions.

Suppose $\sqrt{Q}$ one of the terms of x of order μ; it can occur in x only explicitly, since μ is the maximum order. Since, further, the powers of $\sqrt{Q}$ may be expressed as functions of $\sqrt{Q}$ and Q, which is a term of lower order, we may put

$$x = \frac{a + b\sqrt{Q}}{c + d\sqrt{Q}},$$

where a, b, c, d contain no more than $n - 1$ terms of order μ, besides terms of lower order.

Multiplying both terms of the fraction by $c - d\sqrt{Q}$, $\sqrt{Q}$ disappears from the denominator, and we may write

$$x = \frac{(ac - bdQ) + (bc - ad)\sqrt{Q}}{c^2 - d^2Q} = \alpha + \beta\sqrt{Q},$$

where α and β contain no more than $n - 1$ terms of order μ.

For a second term of order μ we have, in a similar manner, $x = \alpha_1 + \beta_1\sqrt{Q_1}$, etc.

The x *may, therefore, be transformed so as to contain a term of given order* μ *only in its numerator and there only linearly.*

We observe, however, that products of terms of order μ may occur, for α and β still depend upon $n-1$ terms of order μ. We may, then, put

$$\alpha = \alpha_{11} + \alpha_{12}\sqrt{Q_1}, \qquad \beta = \beta_{11} + \beta_{12}\sqrt{Q_1},$$

and hence

$$x = (\alpha_{11} + \alpha_{12}\sqrt{Q_1}) + (\beta_{11} + \beta_{12}\sqrt{Q_1})\sqrt{Q}.$$

6. We proceed in a similar way with the different terms of order $\mu - 1$, which occur explicitly and in Q, Q_1, etc., so that each of these quantities becomes an integral linear function of the term of order $\mu - 1$ under consideration. We then pass on to terms of lower order and finally obtain x, or rather its terms of different orders, under the form of rational integral linear functions of the individual radical expressions which occur explicitly. We then say that x is reduced to the *normal form.*

7. Let m be the total number of independent (**4**) square roots occurring in this normal form. Giving the double sign to these square roots and combining them in all possible ways, we obtain a system of 2^m values

$$x_1, x_2, \ldots \, x_{2^m},$$

which we shall call *conjugate* values.

We must now investigate the equation admitting these conjugate values as roots.

8. These values are not necessarily all distinct; thus, if we have

$$x = \sqrt{a + \sqrt{b}} + \sqrt{a - \sqrt{b}},$$

this expression is not changed when we change the sign of $\sqrt{b}$.

9. If x is an arbitrary quantity and we form the polynomial

$$F(x) = (x - x_1)(x - x_2) \ldots (x - x_{2^m}),$$

$F(x) = 0$ is clearly an equation having as roots these conjugate values. It is of degree 2^m, but may have equal roots (**8**).

The coefficients of the polynomial $F(x)$ *arranged with respect to* x *are rational.*

For let us change the sign of one of the square roots; this will permute two roots, say x_λ and $x_{\lambda'}$, since the roots of $F(x) = 0$ are precisely all the conjugate values. As these roots enter $F(x)$ only under the form of the product

$$(x - x_\lambda)(x - x_{\lambda'}),$$

we merely change the order of the factors of $F(x)$. Hence the polynomial is not changed.

$F(x)$ remains, then, invariable when we change the sign of any one of the square roots; it therefore contains only their squares; and hence $F(x)$ has only rational coefficients.

10. *When any one of the conjugate values satisfies a given equation with rational coefficients,* $f(x) = 0$, *the same is true of all the others.*

$f(x)$ is not necessarily equal to $F(x)$, and may admit other roots besides the x_i's.

Let $x_1 = \alpha + \beta\sqrt{Q}$ be one of the conjugate values; $\sqrt{Q}$, a term of order μ; α and β now depend only upon other terms of order μ and terms of lower order. There must, then, be a conjugate value

$$x_1' = \alpha - \beta\sqrt{Q}.$$

Let us now form the equation $f(x_1) = 0$. $f(x_1)$ may be put into the normal form with respect to $\sqrt{Q}$,

$$f(x_1) = A + B\sqrt{Q}:$$

this expression can equal zero only when A and B are simultaneously zero. Otherwise we should have

$$\sqrt{Q} = -\frac{A}{B};$$

i.e., $\sqrt{Q}$ could be expressed rationally as a function of terms of order μ and of terms of lower order contained in A and B, which is contrary to the hypothesis of the independence of all the square roots (**4**).

But we evidently have

$$f(x_1') = A - B\sqrt{Q};$$

hence if $f(x_1) = 0$, so also $f(x_1') = 0$. Whence the following proposition:

If x_1 *satisfies the equation* $f(x) = 0$, *the same is true of all the conjugate values derived from* x_1 *by changing the signs of the roots of order* μ.

The proof for the other conjugate values is obtained in an analogous manner. Suppose, for example, as may be done without affecting the generality of the reasoning, that the expression x_1 depends on only two terms of order μ, $\sqrt{Q}$ and $\sqrt{Q'}$. $f(x_1)$ may be reduced to the following normal form:

$$(a) \qquad f(x_1) = p + q\sqrt{Q} + r\sqrt{Q'} + s\sqrt{Q}\cdot\sqrt{Q'} = 0.$$

If x_1 depended on more than two terms of order μ, we should only have to add to the preceding expression a greater number of terms of analogous structure.

Equation (*a*) is possible only when we have separately

$$(b) \qquad p = 0, \quad q = 0, \quad r = 0, \quad s = 0.$$

Otherwise $\sqrt{Q}$ and $\sqrt{Q'}$ would be connected by a rational relation, contrary to our hypothesis.

Let now $\sqrt{R}$, $\sqrt{R'}$, ... be the terms of order $\mu - 1$ on which x_1 depends; they occur in p, q, r, s; then can the quantities p, q, r, s, in which they occur, be reduced to the

normal form with respect to $\sqrt{R}$ and $\sqrt{R'}$; and if, for the sake of simplicity, we take only two quantities, $\sqrt{R}$ and $\sqrt{R'}$, we have

$$(c)\qquad p = \kappa_1 + \lambda_1\sqrt{R} + \mu_1\sqrt{R'} + \nu_1\sqrt{R}\,.\,\sqrt{R'} = 0,$$

and three analogous equations for q, r, s.

The hypothesis, already used several times, of the independence of the roots, furnishes the equations

$$(d)\qquad \kappa = 0,\quad \lambda = 0,\quad \mu = 0,\quad \nu = 0.$$

Hence equations (*c*) and consequently $f(x) = 0$ are satisfied when for x_1 we substitute the conjugate values deduced by changing the signs of $\sqrt{R}$ and $\sqrt{R'}$.

Therefore *the equation* $f(x) = 0$ *is also satisfied by all the conjugate values deduced from* x_1 *by changing the signs of the roots of order* $\mu - 1$.

The same reasoning is applicable to the terms of order $\mu - 2$, $\mu - 3$, . . . and our theorem is completely proved.

11. We have so far considered two equations

$$F(x) = 0 \quad\text{and}\quad f(x) = 0.$$

Both have rational coefficients and contain the x_i's as roots. $F(x)$ is of degree 2^m and may have multiple roots; $f(x)$ may have other roots besides the x_i's. We now introduce a third equation, $\phi(x) = 0$, defined as the equation of lowest degree, with rational coefficients, admitting the root x_1 and consequently all the x_i's (**10**).

12. Properties of the Equation $\phi(x) = 0$.

I. $\phi(x) = 0$ *is an irreducible equation, i.e.,* $\phi(x)$ cannot be resolved into two rational polynomial factors. This irreducibility is due to the hypothesis that $\phi(x) = 0$ is the rational equation of *lowest* degree satisfied by the x_i's.

For if we had

$$\phi(x) = \psi(x)\,\chi(x),$$

then $\phi(x_1) = 0$ would require either $\psi(x_1) = 0$, or $\chi(x_1) = 0$, or both. But since these equations are satisfied by all the conjugate values (**10**), $\phi(x) = 0$ would not then be the equation of lowest degree satisfied by the x_i's.

II. $\phi(x) = 0$ *has no multiple roots.* Otherwise $\phi(x)$ could be decomposed into rational factors by the well-known methods of Algebra, and $\phi(x) = 0$ would not be irreducible.

III. $\phi(x) = 0$ *has no other roots than the* x_i's. Otherwise $F(x)$ and $\phi(x)$ would admit a highest common divisor, which could be determined rationally. We could then decompose $\phi(x)$ into rational factors, and $\phi(x)$ would not be irreducible.

IV. Let M be the number of x_i's which have distinct values, and let

$$x_1, x_2, \ldots x_M$$

be these quantities. We shall then have

$$\phi(x) = C(x - x_1)(x - x_2) \ldots (x - x_M).$$

For $\phi(x) = 0$ is satisfied by the quantities x_i and it has no multiple roots. The polynomial $\phi(x)$ is then determined save for a constant factor whose value has no effect upon $\phi(x) = 0$

V. $\phi(x) = 0$ *is the only irreducible equation with rational coefficients satisfied by the* x_i's. For if $f(x) = 0$ were another rational irreducible equation satisfied by x_1 and consequently by the x_i's, $f(x)$ would be divisible by $\phi(x)$ and therefore would not be irreducible.

By reason of the five properties of $\phi(x) = 0$ thus established, we may designate this equation, in short, as *the* irreducible equation satisfied by the x_i's.

13. Let us now compare $F(x)$ and $\phi(x)$. These two polynomials have the x_i's as their only roots, and $\phi(x)$ has no multiple roots. $F(x)$ is, then, divisible by $\phi(x)$; that is,

$$F(x) = F_1(x)\,\phi(x).$$

$F_1(x)$ necessarily has rational coefficients, since it is the quotient obtained by dividing $F(x)$ by $\phi(x)$. If $F_1(x)$ is not a constant it admits roots belonging to $F(x)$; and admitting one it admits all the x_i's (**10**). Hence $F_1(x)$ is also divisible by $\phi(x)$, and

$$F_1(x) = F_2(x)\,\phi(x).$$

If $F_2(x)$ is not a constant the same reasoning still holds, the degree of the quotient being lowered by each operation. Hence at the end of a limited number of divisions we reach an equation of the form

$$F_{\nu-1}(x) = C_1 \cdot \phi(x),$$

and for $F(x)$,

$$F(x) = C_1 \cdot [\phi(x)]^\nu.$$

The polynomial $F(x)$ is then a power of the polynomial of minimum degree $\phi(x)$, except for a constant factor.

14. We can now determine the degree M of $\phi(x)$. $F(x)$ is of degree 2^m; further, it is the νth power of $\phi(x)$. Hence

$$2^m = \nu \cdot M.$$

Therefore M is also a power of 2 and we obtain the following theorem:

The degree of the irreducible equation satisfied by an expression composed of square roots only is always a power of 2.

15. Since, on the other hand, there is only one irreducible equation satisfied by all the x_i's (**12**, V.), we have the converse theorem:

If an irreducible equation is not of degree 2^h, it cannot be solved by square roots.

CHAPTER II.

The Delian Problem and the Trisection of the Angle.

1. Let us now apply the general theorem of the preceding chapter to the *Delian problem*, *i.e.*, to the problem of the *duplication of the cube.* The equation of the problem is manifestly

$$x^3 = 2.$$

This is irreducible, since otherwise $\sqrt[3]{2}$ would have a rational value. For an equation of the third degree which is reducible must have a rational linear factor. Further, the degree of the equation is not of the form 2^h; hence it cannot be solved by means of square roots, and the geometric construction with straight edge and compasses is impossible.

2. Next let us consider the more general equation

$$x^3 = \lambda,$$

λ designating a parameter which may be a complex quantity of the form $a + ib$. This equation furnishes us the analytical expressions for the geometrical problems of the multiplication of the cube and the trisection of an arbitrary angle. The question arises whether this equation is reducible, *i.e.*, whether one of its roots can be expressed as a rational function of λ. It should be remarked that the irreducibility of an expression always depends upon the values of the quantities supposed to be known. In the case $x^3 = 2$, we were dealing with numerical quantities, and the question was whether $\sqrt[3]{2}$ could have a rational numerical value. In the equation $x^3 = \lambda$ we ask whether a root can be represented by a rational function of λ. In the first case, the so-called

domain of rationality comprehends the totality of rational numbers; in the second, it is made up of the rational functions of a parameter. If no limitation is placed upon this parameter we see at once that no expression of the form $\frac{\phi(\lambda)}{\psi(\lambda)}$, in which $\phi(\lambda)$ and $\psi(\lambda)$ are polynomials, can satisfy our equation. Under our hypothesis the equation is therefore irreducible, and since its degree is not of the form 2^h, it cannot be solved by square roots.

3. Let us now restrict the variability of λ. Assume

$$\lambda = r(\cos\phi + i\sin\phi);$$

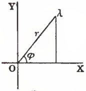


FIG. 1.

whence $\sqrt[3]{\lambda} = \sqrt[3]{r}\;\sqrt[3]{\cos\phi + i\sin\phi}$.

Our problem resolves itself into two, to extract the cube root of a real number and also that of a complex number of the form $\cos\phi + i\sin\phi$, both numbers being regarded as arbitrary. We shall treat these separately.

I. The roots of the equation $x^3 = r$ are

$$\sqrt[3]{r},\ \epsilon\sqrt[3]{r},\ \epsilon^2\sqrt[3]{r},$$

representing by ϵ and ϵ^2 the complex cube roots of unity

$$\epsilon = \frac{-1 + i\sqrt{3}}{2},\quad \epsilon^2 = \frac{-1 - i\sqrt{3}}{2}.$$

Taking for the domain of rationality the totality of rational functions of r, we know by the previous reasoning that the equation $x^3 = r$ is irreducible. Hence the problem of the multiplication of the cube does not admit, in general, of a construction by means of straight edge and compasses.

II. The roots of the equation

$$x^3 = \cos\phi + i\sin\phi$$

are, by De Moivre's formula,

$$x_1 = \cos\frac{\phi}{3} + i\sin\frac{\phi}{3},$$

$$x_2 = \cos\frac{\phi+2\pi}{3} + i\sin\frac{\phi+2\pi}{3},$$

$$x_3 = \cos\frac{\phi+4\pi}{3} + i\sin\frac{\phi+4\pi}{3}.$$

These roots are represented geometrically by the vertices of an equilateral triangle inscribed in the circle with radius unity and center at the origin. The figure shows that to the root x_1 corresponds the argument $\frac{\phi}{3}$. Hence the equation

$$x^3 = \cos\phi + i\sin\phi$$

is the analytic expression of the problem of the trisection of the angle.

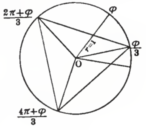


FIG. 2

If this equation were reducible, one, at least, of its roots could be represented as a rational function of $\cos\phi$ and $\sin\phi$, its value remaining unchanged on substituting $\phi + 2\pi$ for ϕ. But if we effect this change by a continuous variation of the angle ϕ, we see that the roots x_1, x_2, x_3 undergo a cyclic permutation. *Hence no root can be represented as a rational function of* $\cos\phi$ *and* $\sin\phi$. The equation under consideration is irreducible and *therefore cannot be solved by the aid of a finite number of square roots.* Hence *the trisection of the angle cannot be effected with straight edge and compasses.*

This demonstration and the general theorem evidently hold good only when ϕ is an arbitrary angle; but for certain special values of ϕ the construction may prove to be possible, *e.g.*, when $\phi = \frac{\pi}{2}$.

CHAPTER III.

The Division of the Circle into Equal Parts.

1. The problem of dividing a given circle into n equal parts has come down from antiquity; for a long time we have known the possibility of solving it when $n = 2^h$, 3, 5, or the product of any two or three of these numbers. In his *Disquisitiones Arithmeticae,* Gauss extended this series of numbers by showing that the division is possible for every prime number of the form $p = 2^{2^\mu} + 1$ but impossible for all other prime numbers and their powers. If in $p = 2^{2^\mu} + 1$ we make $\mu = 0$ and 1, we get $p = 3$ and 5, cases already known to the ancients. For $\mu = 2$ we get $p = 2^{2^2} + 1 = 17$, a case completely discussed by Gauss.

For $\mu = 3$ we get $p = 2^{2^3} + 1 = 257$, likewise a prime number. The regular polygon of 257 sides can be constructed. Similarly for $\mu = 4$, since $2^{2^4} + 1 = 65537$ is a prime number. $\mu = 5$, $\mu = 6$, $\mu = 7$, $\mu = 8$, $\mu = 9$, $\mu = 11$, $\mu = 12$, $\mu = 15$, $\mu = 18$, $\mu = 23$, $\mu = 36$, $\mu = 38$, $\mu = 73$ give no prime numbers. The proof that the large numbers corresponding to $\mu = 5, 6, \ldots, 73$ are not prime has required a large expenditure of labor and ingenuity. It is, therefore, quite possible that $\mu = 4$ is the last number for which a solution can be effected.

Upon the regular polygon of 257 sides Richelot published an extended investigation in Crelle's *Journal,* IX, 1832, pp. 1–26, 146–161, 209–230, 337–356. The title of the memoir is . *De resolutione algebraica aequationis* $x^{257} = 1$, *sive de divisione circuli per bisectionem anguli septies repetitam in partes* 257 *inter se aequales commentatio coronata.*

To the regular polygon of 65537 sides Professor Hermes of Lingen devoted ten years of his life, examining with care all the roots furnished by Gauss's method. His MSS. are preserved in the collection of the mathematical seminary in Göttingen. (Compare a communication of Professor Hermes in No. 3 of the *Göttinger Nachrichten* for 1894.)

2. We may restrict the problem of the division of the circle into n equal parts to the cases where n is a prime number p or a power p^a of such a number. For if n is a composite number and if μ and ν are factors of n, prime to each other, we can always find integers a and b, positive or negative, such that

$$1 = a\mu + b\nu;$$

whence

$$\frac{1}{\mu\nu} = \frac{a}{\nu} + \frac{b}{\mu}.$$

To divide the circle into $\mu\nu = n$ equal parts it is sufficient to know how to divide it into μ and ν equal parts respectively. Thus, for $n = 15$, we have

$$\frac{1}{15} = \frac{2}{3} - \frac{3}{5}.$$

3. As will appear, the division into p equal parts (p being a prime number) is possible only when p is of the form $p = 2^h + 1$. We shall next show that a prime number can be of this form only when $h = 2^\mu$. For this we shall make use of Fermat's Theorem :

If p *is a prime number and* a *an integer not divisible by* p, *these numbers satisfy the congruence*

$$a^{p-1} \equiv +1 \pmod{p}.$$

$p-1$ is not necessarily the lowest exponent which, for a given value of a, satisfies the congruence. If s is the lowest exponent it may be shown that s is a divisor of $p-1$. In particular, if $s = p-1$ we say that a is a *primitive root* of p,

and notice that for every prime number p there is a primitive root. We shall make use of this notion further on.

Suppose, then, p a prime number such that

$$(1) \qquad p = 2^h + 1,$$

and s the least integer satisfying

$$(2) \qquad 2^s \equiv +1 \pmod{p}.$$

From (1) $2^h < p$; from (2) $2^s > p$.

$$\therefore s > h.$$

(1) shows that h is the least integer satisfying the congruence

$$(3) \qquad 2^h \equiv -1 \pmod{p}.$$

From (2) and (3), by division,

$$2^{s-h} \equiv -1 \pmod{p}.$$

$$\therefore (4) \qquad s - h \not< h, \quad s \not< 2h.$$

From (3), by squaring,

$$2^{2h} \equiv 1 \pmod{p}.$$

Comparing with (2) and observing that s is the least exponent satisfying congruences of the form

$$2^x \equiv 1 \pmod{p},$$

we have

$$(5) \qquad s \not> 2h.$$

$$\therefore s = 2h.$$

We have observed that s is a divisor of $p - 1 = 2^h$; the same is true of h, which is, therefore, a power of 2. Hence prime numbers of the form $2^h + 1$ are necessarily of the form $2^{2^\mu} + 1$.

4. This conclusion may be established otherwise. Suppose that h is divisible by an odd number, so that

$$h = h'(2n + 1);$$

then, by reason of the formula

$$x^{2n+1} + 1 = (x + 1)(x^{2n} - x^{2n-1} + \ldots - x + 1),$$

$p = 2^{h \cdot (2n+1)} + 1$ is divisible by $2^h + 1$, and hence is not a prime number.

5. We now reach our fundamental proposition :

p *being a prime number, the division of the circle into* p *equal parts by the straight edge and compasses is impossible unless* p *is of the form*

$$p = 2^h + 1 = 2^{2^\mu} + 1.$$

Let us trace in the z-plane $(z = x + iy)$ a circle of radius 1. To divide this circle into n equal parts, beginning at $z = 1$, is the same as to solve the equation

$$z^n - 1 = 0.$$

This equation admits the root $z = 1$; let us suppress this root by dividing by $z - 1$, which is the same geometrically as to disregard the initial point of the division. We thus obtain the equation

$$z^{n-1} + z^{n-2} + \ldots + z + 1 = 0,$$

which may be called the *cyclotomic equation.* As noticed above, we may confine our attention to the cases where n is a prime number or a power of a prime number. We shall first investigate the case when $n = p$. The essential point of the proof is to show *that the above equation is irreducible.* For since, as we have seen, irreducible equations can only be solved by means of square roots in finite number when their degree is a power of 2, a division into p parts is always impossible when $p - 1$ is not equal to a power of 2, *i.e.*, when

$$p \neq 2^h + 1 \neq 2^{2^\mu} + 1.$$

Thus we see why Gauss's prime numbers occupy such an exceptional position.

6. At this point we introduce a lemma known as *Gauss's Lemma.* If

$$F(z) = z^m + Az^{m-1} + Bz^{m-2} + \ldots + Lz + M,$$

where A, B, . . . are integers, and $F(z)$ can be resolved into two rational factors $f(z)$ and $\phi(z)$, so that

$$F(z) = f(z) \cdot \phi(z) = (z^{m'} + \alpha_1 z^{m'-1} + \alpha_2 z^{m'-2} + \ldots)$$
$$(z^{m''} + \beta_1 z^{m''-1} + \beta_2 z^{m''-2} + \ldots),$$

then must the α's and β's also be integers. In other words :

If an integral expression can be resolved into rational factors these factors must be integral expressions.

Let us suppose the α's and β's to be fractional. In each factor reduce all the coefficients to the least common denominator. Let a_0 and b_0 be these common denominators. Finally multiply both members of our equation by $a_0 b_0$. It takes the form

$$a_0 b_0 F(z) = f_1(z)\, \phi_1(z) = (a_0 z^{m'} + a_1 z^{m'-1} + \ldots)$$
$$(b_0 z^{m''} + b_1 z^{m''-1} + \ldots).$$

The a's are integral and prime to one another, as also the b's, since a_0 and b_0 are the least common denominators.

Suppose a_0 and b_0 different from unity and let q be a prime divisor of $a_0 b_0$. Further, let a_i be the first coefficient of $f_1(z)$ and b_k the first coefficient of $\phi_1(z)$ not divisible by q. Let us develop the product $f_1(z)\, \phi_1(z)$ and consider the coefficient of $z^{m'+m''-i-k}$. It will be

$$a_i b_k + a_{i-1} b_{k+1} + a_{i-2} b_{k+2} + \ldots + a_{i+1} b_{k-1} + a_{i+2} b_{k-2} + \ldots$$

According to our hypotheses, all the terms after the first are divisible by q, but the first is not. Hence this coefficient is not divisible by q. Now the coefficient of $z^{m'+m''-i-k}$ in the first member is divisible by $a_0 b_0$, *i.e.*, by q. Hence if the identity is true it is impossible for a coefficient not divisible by q to occur in each polynomial. The coefficients of one at least of the polynomials are then all divisible by q. Here is another absurdity, since we have seen that all the coefficients are

prime to one another. Hence we cannot suppose a_0 and b_0 different from 1, and consequently the α's and β's are integral.

7. In order to show that the cyclotomic equation is irreducible, it is sufficient to show by Gauss's Lemma that the first member cannot be resolved into factors with integral coefficients. To this end we shall employ the simple method due to Eisenstein, in Crelle's *Journal*, XXXIX, p. 167, which depends upon the substitution

$$z = x + 1.$$

We obtain

$$f(z) = \frac{z^p - 1}{z - 1} = \frac{(x+1)^p - 1}{x} = x^{p-1} + px^{p-2} + \frac{p(p-1)}{1 \cdot 2} x^{p-3} + \ldots + \frac{p(p-1)}{1 \cdot 2} x + p = 0.$$

All the coefficients of the expanded member except the first are divisible by p; the last coefficient is always p itself, by hypothesis a prime number. An expression of this class is always irreducible.

For if this were not the case we should have

$$f(x+1) = (x^m + a_1 x^{m-1} + \ldots + a_{m-1} x + a_m)(x^{m'} + b_1 x^{m'-1} + \ldots + b_{m'-1} x + b_{m'}),$$

where the a's and b's are integers.

Since the term of zero degree in the above expression of $f(z)$ is p, we have $a_m b_{m'} = p$. p being prime, one of the factors of $a_m b_{m'}$ must be unity. Suppose, then,

$$a_m = \pm p, \quad b_{m'} = \pm 1.$$

Equating the coefficients of the terms in x, we have

$$\frac{p(p-1)}{2} = a_{m-1} b_{m'} + a_m b_{m'-1}.$$

The first member and the second term of the second being divisible by p, $a_{m-1}b_m$ must be so also. Since $b_m = \pm 1$, a_{m-1} is divisible by p. Equating the coefficients of the terms in x^2 we may show that a_{m-2} is divisible by p. Similarly we show that all of the remaining coefficients of the factor $x^m + a_1x^{m-1} + \ldots + a_{m-1}x + a_m$ are divisible by p. But this cannot be true of the coefficient of x^m, which is 1. The assumed equality is impossible and hence the cyclotomic equation is irreducible when p is a prime.

8. We now consider the case where n is a power of a prime number, say $n = p^\alpha$. We propose to show that when $p > 2$ the division of the circle into p^2 equal parts is impossible. The general problem will then be solved, since the division into p^α equal parts evidently includes the division into p^2 equal parts.

The cyclotomic equation is now

$$\frac{z^{p^2} - 1}{z - 1} = 0.$$

It admits as roots extraneous to the problem those which come from the division into p equal parts, *i.e.*, the roots of the equation

$$\frac{z^p - 1}{z - 1} = 0.$$

Suppressing these roots by division we obtain

$$f(z) = \frac{z^{p^2} - 1}{z^p - 1} = 0$$

as the cyclotomic equation. This may be written

$$z^{p(p-1)} + z^{p(p-2)} + \ldots + z^p + 1 = 0.$$

Transforming by the substitution

$$z = x + 1,$$

we have

$$(x+1)^{p(p-1)} + (x+1)^{p(p-2)} + \ldots + (x+1)^p + 1 = 0.$$

The number of terms being p, the term independent of x after development will be equal to p, and the sum will take the form

$$x^{p(p-1)} + p \cdot \chi(x),$$

where $\chi(x)$ is a polynomial with integral coefficients whose constant term is 1. We have just shown that such an expression is always irreducible. Consequently *the new cyclotomic equation is also irreducible.*

The degree of this equation is $p(p-1)$. On the other hand an irreducible equation is solvable by square roots only when its degree is a power of 2. Hence a circle is divisible into p^2 equal parts only when $p = 2$, p being assumed to be a prime.

The same is true, as already noted, for the division into p^α equal parts when $\alpha > 2$.

CHAPTER IV.

The Construction of the Regular Polygon of 17 Sides.

1. We have just seen that the division of the circle into equal parts by the straight edge and compasses is possible only for the prime numbers studied by Gauss. It will now be of interest to learn how the construction can actually be effected.

The purpose of this chapter, then, will be to show in an elementary way how to inscribe in the circle the regular polygon of 17 sides.

Since we possess as yet no method of construction based upon considerations purely geometrical, we must follow the path indicated by our general discussions. We consider, first of all, the roots of the cyclotomic equation

$$x^{16} + x^{15} + \ldots + x^2 + x + 1 = 0,$$

and construct geometrically the expression, formed of square roots, deduced from it.

We know that the roots can be put into the transcendental form

$$\epsilon_\kappa = \cos\frac{2\kappa\pi}{17} + i\sin\frac{2\kappa\pi}{17} \quad (\kappa = 1, 2, \ldots 16);$$

and if

$$\epsilon_1 = \cos\frac{2\pi}{17} + i\sin\frac{2\pi}{17},$$

that

$$\epsilon_\kappa = \epsilon_1^{\kappa}.$$

Geometrically, these roots are represented in the complex plane by the vertices, different from 1, of the regular polygon of 17 sides inscribed in a circle of radius 1, having the origin

as center. The selection of ϵ_1 is arbitrary, but for the construction it is essential to indicate some ϵ as the point of departure. Having fixed upon ϵ_1, the angle corresponding to ϵ_κ is κ times the angle corresponding to ϵ_1, which completely determines ϵ_κ.

2. The fundamental idea of the solution is the following: *Forming a primitive root to the modulus* 17 *we may arrange the* 16 *roots of the equation in a cycle in a determinate order.*

As already stated, a number a is said to be a primitive root to the modulus 17 when the congruence

$$a^s \equiv +1 \pmod{17}$$

has for least solution $s = 17 - 1 = 16$. The number 3 possesses this property; for we have

$$\left.\begin{array}{llll} 3^1 \equiv 3 & 3^5 \equiv 5 & 3^9 \equiv 14 & 3^{13} \equiv 12 \\ 3^2 \equiv 9 & 3^6 \equiv 15 & 3^{10} \equiv 8 & 3^{14} \equiv 2 \\ 3^3 \equiv 10 & 3^7 \equiv 11 & 3^{11} \equiv 7 & 3^{15} \equiv 6 \\ 3^4 \equiv 13 & 3^8 \equiv 16 & 3^{12} \equiv 4 & 3^{16} \equiv 1 \end{array}\right\} \pmod{17}.$$

Let us then arrange the roots ϵ_κ so that their subscripts are the preceding remainders in order

$$\epsilon_3, \epsilon_9, \epsilon_{10}, \epsilon_{13}, \epsilon_5, \epsilon_{15}, \epsilon_{11}, \epsilon_{16}, \epsilon_{14}, \epsilon_8, \epsilon_7, \epsilon_4, \epsilon_{12}, \epsilon_2, \epsilon_6, \epsilon_1.$$

Notice that if r is the remainder of $3^\kappa \pmod{17}$, we have

$$3^\kappa = 17q + r,$$

whence

$$\epsilon_r = \epsilon_1{}^r = \epsilon_1{}^{3^\kappa}$$

If r′ is the next remainder, we have similarly

$$\epsilon_{r'} = \epsilon_1{}^{3^{\kappa+1}} = (\epsilon_1{}^{3^\kappa})^3 = (\epsilon_r)^3.$$

Hence in this series of roots each root is the cube of the preceding.

Gauss's method consists in decomposing this cycle into sums containing 8, 4, 2, 1 roots respectively, corresponding to the divisors of 16. Each of these sums is called a period.

The periods thus obtained may be calculated successively as roots of certain quadratic equations.

The process just outlined is only a particular case of that employed in the general case of the division into p equal parts. The $p-1$ roots of the cyclotomic equation are cyclically arranged by means of a primitive root of p, and the periods may be calculated as roots of certain auxiliary equations. The degree of these last depends upon the prime factors of $p-1$. They are not necessarily equations of the second degree.

The general case has, of course, been treated in detail by Gauss in his *Disquisitiones*, and also by Bachmann in his work, *Die Lehre von der Kreisteilung* (Leipzig, 1872).

3. In our case of the 16 roots the periods may be formed in the following manner: Form two periods of 8 roots by taking in the cycle, first, the roots of even order, then those of odd order. Designate these periods by x_1 and x_2, and replace each root by its index. We may then write symbolically

$$x_1 = 9 + 13 + 15 + 16 + 8 + 4 + 2 + 1,$$
$$x_2 = 3 + 10 + 5 + 11 + 14 + 7 + 12 + 6.$$

Operating upon x_1 and x_2 in the same way, we form 4 periods of 4 terms:

$$y_1 = 13 + 16 + 4 + 1,$$
$$y_2 = 9 + 15 + 8 + 2,$$
$$y_3 = 10 + 11 + 7 + 6,$$
$$y_4 = 3 + 5 + 14 + 12.$$

Operating in the same way upon the y's, we obtain 8 periods of 2 terms:

$$z_1 = 16 + 1, \qquad z_5 = 11 + 6,$$
$$z_2 = 13 + 4, \qquad z_6 = 10 + 7,$$
$$z_3 = 15 + 2, \qquad z_7 = 5 + 12,$$
$$z_4 = 9 + 8, \qquad z_8 = 3 + 14.$$

It now remains to show that *these periods can be calculated successively by the aid of square roots.*

4. It is readily seen that the sum of the remainders corresponding to the roots forming a period z is always equal to 17. These roots are then ϵ_r and ϵ_{17-r};

$$\epsilon_r = \cos r\frac{2\pi}{17} + i \sin r\frac{2\pi}{17},$$

$$\epsilon_{r'} = \epsilon_{17-r} = \cos (17-r)\frac{2\pi}{17} + i \sin (17-r)\frac{2\pi}{17},$$

$$= \cos r\frac{2\pi}{17} - i \sin r\frac{2\pi}{17}.$$

Hence

$$\epsilon_r + \epsilon_{r'} = 2 \cos r\frac{2\pi}{17}.$$

Therefore all the periods z are real, and we readily obtain

$$z_1 = 2 \cos \frac{2\pi}{17}, \qquad z_5 = 2 \cos 6\frac{2\pi}{17},$$

$$z_2 = 2 \cos 4\frac{2\pi}{17}, \qquad z_6 = 2 \cos 7\frac{2\pi}{17},$$

$$z_3 = 2 \cos 2\frac{2\pi}{17}, \qquad z_7 = 2 \cos 5\frac{2\pi}{17},$$

$$z_4 = 2 \cos 8\frac{2\pi}{17}, \qquad z_8 = 2 \cos 3\frac{2\pi}{17}.$$

Moreover, by definition,

$$x_1 = z_1 + z_2 + z_3 + z_4, \qquad x_2 = z_5 + z_6 + z_7 + z_8,$$

$$y_1 = z_1 + z_2, \quad y_2 = z_3 + z_4, \quad y_3 = z_5 + z_6, \quad y_4 = z_7 + z_8.$$

5. It will be necessary to determine the relative magnitude of the different periods. For this purpose we shall employ the following artifice: We divide the semicircle of unit radius into 17 equal parts and denote by S_1, S_2, . . . S_{17} the distances

of the consecutive points of division $A_1, A_2, \ldots A_{17}$ from the initial point of the semicircle, S_{17} being equal to the diameter, *i.e.*, equal to 2. The angle $A_\kappa A_{17} O$ has the same measure as the half of the arc $A_\kappa O$, which equals $\frac{2\kappa\pi}{34}$. Hence

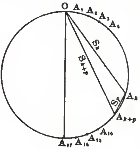


FIG. 3.

$$S_\kappa = 2 \sin \frac{\kappa\pi}{34} = 2 \cos \frac{(17-\kappa)\pi}{34}.$$

That this may be identical with $2 \cos h \frac{2\pi}{17}$, we must have

$$4h = 17 - \kappa,$$
$$\kappa = 17 - 4h.$$

Giving to h the values 1, 2, 3, 4, 5, 6, 7, 8, we find for κ the values 13, 9, 5, 1, -3, -7, -11, -15. Hence

$$z_1 = S_{13}, \qquad z_5 = -S_7,$$
$$z_2 = S_1, \qquad z_6 = -S_{11},$$
$$z_3 = S_9, \qquad z_7 = -S_3,$$
$$z_4 = -S_{15}, \qquad z_8 = S_5.$$

The figure shows that S_κ increases with the subscript; hence the order of increasing magnitude of the periods z is

$$z_4,\ z_6,\ z_5,\ z_7,\ z_2,\ z_8,\ z_3,\ z_1.$$

Moreover, the chord $A_\kappa A_{\kappa+p}$ subtends p divisions of the semicircumference and is equal to S_p; the triangle $OA_\kappa A_{\kappa+p}$ shows that

$$S_{\kappa+p} < S_\kappa + S_p,$$

and *a fortiori*

$$S_{\kappa+p} < S_{\kappa+r} + S_{p+r}.$$

Calculating the differences two and two of the periods y, we easily find

$$\begin{aligned}
y_1 - y_2 &= S_{13} + S_1 \; - S_9 + S_{15} > 0,\\
y_1 - y_3 &= S_{13} + S_1 \; + S_7 + S_{11} > 0,\\
y_1 - y_4 &= S_{13} + S_1 \; + S_3 - S_5 \; > 0,\\
y_2 - y_3 &= S_9 \; - S_{15} + S_7 + S_{11} > 0,\\
y_2 - y_4 &= S_9 \; - S_{15} + S_8 - S_5 \; < 0,\\
y_3 - y_4 &= - S_7 - S_{11} + S_8 - S_5 < 0.
\end{aligned}$$

Hence

$$y_3 < y_2 < y_4 < y_1.$$

Finally we obtain in a similar way

$$x_2 < x_1.$$

6. We now propose to calculate $z_1 = 2 \cos \frac{2\pi}{17}$. After making this calculation and constructing z_1, we can easily deduce the side of the regular polygon of 17 sides. In order to find the quadratic equation satisfied by the periods, we proceed to determine symmetric functions of the periods.

Associating z_1 with the period z_2 and thus forming the period y_1, we have, first,

$$z_1 + z_2 = y_1.$$

Let us now determine $z_1 z_2$. We have

$$z_1 z_2 = (16 + 1)(13 + 4),$$

where the symbolic product κp represents

$$\epsilon_\kappa \cdot \epsilon_p = \epsilon_{\kappa + p}.$$

Hence it should be represented symbolically by $\kappa + p$, remembering to subtract 17 from $\kappa + p$ as often as possible. Thus,

$$z_1 z_2 = 12 + 3 + 14 + 5 = y_4.$$

Therefore z_1 and z_2 are the roots of the quadratic equation

$$(\zeta) \qquad z^2 - y_1 z + y_4 = 0,$$

whence, since $z_1 > z_2$,

$$z_1 = \frac{y_1 + \sqrt{y_1^2 - 4y_4}}{2}, \qquad z_2 = \frac{y_1 - \sqrt{y_1^2 - 4y_4}}{2}.$$

We must now determine y_1 and y_4. Associating y_1 with the period y_2, thus forming the period x_1, and y_3 with the period y_4, thus forming the period x_2, we have, first,

$$y_1 + y_2 = x_1.$$

Then,

$$y_1 y_2 = (13 + 16 + 4 + 1)\,(9 + 15 + 8 + 2).$$

Expanding symbolically, the second member becomes equal to the sum of all the roots; that is, to -1. Therefore y_1 and y_2 are the roots of the equation

$$(\eta) \qquad y^2 - x_1 y - 1 = 0,$$

whence, since $y_1 > y_2$,

$$y_1 = \frac{x_1 + \sqrt{x_1^2 + 4}}{2}, \qquad y_2 = \frac{x_1 - \sqrt{x_1^2 + 4}}{2}$$

Similarly,

$$y_3 + y_4 = x_2$$

and

$$y_3 y_4 = -1.$$

Hence y_3 and y_4 are the roots of the equation

$$(\eta') \qquad y^2 - x_2 y - 1 = 0\,;$$

whence, since $y_4 > y_3$,

$$y_4 = \frac{x_2 + \sqrt{x_2^2 + 4}}{2}, \qquad y_3 = \frac{x_2 - \sqrt{x_2^2 + 4}}{2}$$

It now remains to determine x_1 and x_2. Since $x_1 + x_2$ is equal to the sum of all the roots,

$$x_1 + x_2 = -1.$$

Further,

$$x_1 x_2 = (13 + 16 + 4 + 1 + 9 + 15 + 8 + 2)\,(10 + 11 + 7 + 6 + 3 + 5 + 14 + 12).$$

Expanding symbolically, each root occurs 4 times, and thus

$$x_1 x_2 = -4.$$

Therefore x_1 and x_2 are the roots of the quadratic

$$(\xi) \qquad x^2 + x - 4 = 0;$$

whence, since $x_1 > x_2$,

$$x_1 = \frac{-1+\sqrt{17}}{2}, \qquad x_2 = \frac{-1-\sqrt{17}}{2}$$

Solving equations ξ, η, η', ζ in succession, z_1 is determined by a series of square roots.

Effecting the calculations, we see that z_1 depends upon the four square roots

$$\sqrt{17}, \ \sqrt{x_1^2+4}, \ \sqrt{x_2^2+4}, \ \sqrt{y_1^2-4y_4}.$$

If we wish to reduce z_1 to the normal form we must see whether any one of these square roots can be expressed rationally in terms of the others.

Now, from the roots of (η),

$$\sqrt{x_1^2+4} = y_1 - y_2,$$
$$\sqrt{x_2^2+4} = y_4 - y_3.$$

Expanding symbolically, we verify that

$$(y_1 - y_2)(y_4 - y_3) = 2(x_1 - x_2),^*$$

* $(y_1 - y_2)(y_4 - y_3) = (13 + 16 + 4 + 1 - 9 - 15 - 8 - 2)(3 + 5 + 14 + 12 - 10 - 11 - 7 - 6)$

$$\begin{aligned}
&= 16 + 1 + 10 + 8 - 6 - 7 - 3 - 2 \\
&\quad + 2 + 4 + 13 + 11 - 9 - 10 - 6 - 5 \\
&\quad + 7 + 9 + 1 + 16 - 14 - 15 - 11 - 10 \\
&\quad + 4 + 6 + 15 + 13 - 11 - 12 - 8 - 7 \\
&\quad - 12 - 14 - 6 - 4 + 2 + 3 + 16 + 15 \\
&\quad - 1 - 3 - 12 - 10 + 8 + 9 + 5 + 4 \\
&\quad - 11 - 13 - 5 - 3 + 1 + 2 + 15 + 14 \\
&\quad - 5 - 7 - 16 - 14 + 12 + 13 + 9 + 8
\end{aligned}$$

$= 2(16 + 1 + 8 + 2 + 4 + 13 + 15 + 9 - 10 - 6 - 7 - 3 - 11 - 5 - 14 - 12)$

$= 2(x_1 - x_2).$

that is,

$$\sqrt{x_1^2+4}\ \sqrt{x_2^2+4}=2\sqrt{17}.$$

Hence $\sqrt{x_2^2+4}$ can be expressed rationally in terms of the other two square roots. This equation shows that if two of the three differences y_1-y_2, y_4-y_3, x_1-x_2 are positive, the same is true of the third, which agrees with the results obtained directly.

Replacing now x_1, y_1, y_4 by their numerical values, we obtain in succession

$$x_1=\frac{-1+\sqrt{17}}{2},$$

$$y_1=\frac{-1+\sqrt{17}+\sqrt{34-2\sqrt{17}}}{4},$$

$$y_4=\frac{-1-\sqrt{17}+\sqrt{34+2\sqrt{17}}}{4},$$

$$z_1=\frac{-1+\sqrt{17}+\sqrt{34-2\sqrt{17}}}{8}$$

$$+\frac{\sqrt{68+12\sqrt{17}-16\sqrt{34+2\sqrt{17}}-2(1-\sqrt{17})\sqrt{34-2\sqrt{17}}}}{8}$$

The algebraic part of the solution of our problem is now completed. We have already remarked that there is no known construction of the regular polygon of 17 sides based upon purely geometric considerations. There remains, then, only the geometric translation of the individual algebraic steps.

7. We may be allowed to introduce here a brief historical account of geometric constructions with straight edge and compasses.

In the geometry of the ancients the straight edge and compasses were always used together; the difficulty lay merely in bringing together the different parts of the figure so as not to

draw any unnecessary lines. Whether the several steps in the construction were made with straight edge or with compasses was a matter of indifference.

On the contrary, in 1797, the Italian Mascheroni succeeded in effecting all these constructions with the compasses alone; he set forth his methods in his *Geometria del compasso*, and claimed that constructions with compasses were practically more exact than those with the straight edge. As he expressly stated, he wrote for mechanics, and therefore with a practical end in view. Mascheroni's original work is difficult to read, and we are under obligations to Hutt for furnishing a brief *résumé* in German, *Die Mascheroni'schen Constructionen* (Halle, 1880).

Soon after, the French, especially the disciples of Carnot, the author of the *Géométrie de position*, strove, on the other hand, to effect their constructions as far as possible with the straight edge. (See also Lambert, *Freie Perspective*, 1774.)

Here we may ask a question which algebra enables us to answer immediately: In what cases can the solutiou of an algebraic problem be constructed with the straight edge alone? The answer is not given with sufficient explicituess by the authors mentioned. We shall say:

With the straight edge alone we can construct all algebraic expressions whose form is rational.

With a similar view Brianchon published in 1818 a paper, *Les applications de la théorie des transversales*, in which he shows how his constructions can be effected in many cases with the straight edge alone. He likewise insists upon the practical value of his methods, which are especially adapted to field work in surveying.

Poncelet was the first, in his *Traité des propriétés projectives* (Vol. I, Nos. 351–357), to conceive the idea that it is sufficient to use a *single fixed circle* in connection with the straight lines

of the plane in order to construct all expressions depending upon square roots, the center of the fixed circle being given.

This thought was developed by Steiner in 1833 in a celebrated paper entitled *Die geometrischen Constructionen, ausgeführt mittels der geraden Linie und eines festen Kreises, als Lehrgegenstand für höhere Unterrichtsanstalten und zum Selbstunterricht.*

8. To construct the regular polygon of 17 sides we shall follow the method indicated by von Staudt (Crelle's *Journal,* Vol. XXIV, 1842), modified later by Schröter (Crelle's *Journal,* Vol. LXXV, 1872). The construction of the regular polygon of 17 sides is made in accordance with the methods indicated by Poncelet and Steiner, inasmuch as besides the straight edge but one fixed circle is used.*

First, we will show *how with the straight edge and one fixed circle we can solve every quadratic equation.*

At the extremities of a diameter of the fixed unit circle (Fig. 4) we draw two tangents, and select the lower as the axis of X, and the diameter perpendicular to it as the axis of Y. Then the equation of the circle is

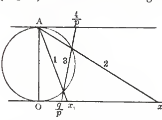


FIG. 4.

$$x^2 + y(y - 2) = 0.$$

Let

$$x^2 - px + q = 0$$

be any quadratic equation with real roots x_1 and x_2. Required to construct the roots x_1 and x_2 upon the axis of X.

Lay off upon the upper tangent from A to the right, a segment measured by $\frac{4}{p}$; upon the axis of X from O, a segment

* A Mascheroni construction of the regular polygon of 17 sides by L Gérard is given in *Math. Annalen,* Vol XLVIII, 1896, pp. 390–392.

measured by $\frac{q}{p}$; connect the extremities of these segments by the line 3 and project the intersections of this line with the circle from A, by the lines 1 and 2, upon the axis of X. The segments thus cut off upon the axis of X are measured by x_1 and x_2.

Proof. Calling the intercepts upon the axis of X, x_1 and x_2, we have the equation of the line 1,

$$2x + x_1(y - 2) = 0\,;$$

of the line 2.

$$2x + x_2(y - 2) = 0.$$

If we multiply the first members of these two equations we get

$$x^2 + \frac{x_1 + x_2}{2} x(y - 2) + \frac{x_1 x_2}{4}(y - 2)^2 = 0$$

as the equation of the line pair formed by 1 and 2. Subtracting from this the equation of the circle, we obtain

$$\frac{x_1 + x_2}{2} x(y - 2) + \frac{x_1 x_2}{4}(y - 2)^2 - y(y - 2) = 0$$

This is the equation of a conic passing through the four intersections of the lines 1 and 2 with the circle. From this equation we can remove the factor $y - 2$, corresponding to the tangent, and we have left

$$\frac{x_1 + x_2}{2} x + \frac{x_1 x_2}{4}(y - 2) - y = 0,$$

which is the equation of the line 3. If we now make $x_1 + x_2 = p$ and $x_1 x_2 = q$, we get

$$\frac{p}{2} x + \frac{q}{4}(y - 2) - y = 0,$$

and the transversal 3 cuts off from the line $y = 2$ the seg-

ment $\frac{4}{p}$, and from the line $y=0$ the segment $\frac{q}{p}$. Thus the correctness of the construction is established.

9. In accordance with the method just explained, we shall now construct the roots of our four quadratic equations. They are (see pp. 29–31)

$(\xi)\quad x^2 + \ \ x - 4 \ = 0$, with roots x_1 and x_2; $x_1 > x_2$,
$(\eta)\quad y^2 - x_1 y - 1 \ = 0$, with roots y_1 and y_2; $y_1 > y_2$,
$(\eta')\quad y^2 - x_2 y - 1 \ = 0$, with roots y_3 and y_4; $y_4 > y_3$,
$(\zeta)\quad z^2 - y_1 z + y_4 = 0$, with roots z_1 and z_2; $z_1 > z_2$.

These will furnish

$$z_1 = 2 \cos \frac{2\pi}{17},$$

whence it is easy to construct the polygon desired. We notice further that to construct z_1 it is sufficient to construct x_1, x_2, y_1, y_4.

We then lay off the following segments: upon the upper tangent, $y=2$,

$$-4, \ \frac{4}{x_1}, \ \frac{4}{x_2}, \ \frac{4}{y_1};$$

upon the axis of X,

$$+4, \ -\frac{1}{x_1}, \ -\frac{1}{x_2}, \ \frac{y_4}{y_1}.$$

This may all be done in the following manner: The straight line connecting the point $+4$ upon the axis of X with the point -4 upon the tangent $y=2$ cuts the circle in

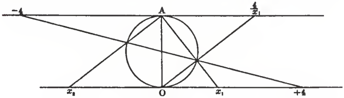


FIG. 5.

two points, the projection of which from the point A $(0, 2)$, the upper vertex of the circle, gives the two roots x_1, x_2 of the first quadratic equation as intercepts upon the axis of X.

To solve the second equation we have to lay off $\frac{4}{x_1}$ above and $-\frac{1}{x_1}$ below.

To determine the first point we connect x_1 upon the axis of X with A, the upper vertex, and from O, the lower vertex, draw another straight line through the intersection of this line with the circle. This cuts off upon the upper tangent the intercept $\frac{4}{x_1}$. This can easily be shown analytically.

The equation of the line from A to x_1 (Fig. 5),

$$2x + x_1y = 2x_1,$$

and that of the circle,

$$x^2 + y(y-2) = 0,$$

give as the coördinates of their intersection

$$\frac{4x_1}{x_1^2+4},\ \frac{2x_1^2}{x_1^2+4}.$$

The equation of the line from O through this point becomes

$$y = \frac{x_1}{2}x,$$

cutting off upon $y = 2$ the intercept $\frac{4}{x_1}$.

We reach the same conclusion still more simply by the use of some elementary notions of projective geometry. By our construction we have obviously associated with every point x of the lower range one, and only one, point of the upper, so that to the point $x = \infty$ corresponds the point $x' = 0$, and conversely. Since in such a correspondence there must exist a

linear relation, the abscissa x' of the upper point must satisfy the equation

$$x' = \frac{\text{const.}}{x}$$

Since $x' = 2$ when $x = 2$, as is obvious from the figure, the constant $= 4$.

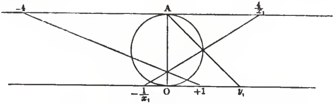


FIG. 6.

To determine $-\frac{1}{x_1}$ upon the axis of X we connect the point -4 upon the upper with the point $+1$ upon the lower tangent (Fig. 6). The point thus determined upon the vertical diameter we connect with the point $\frac{4}{x_1}$ above. This line cuts off upon the axis of X the intercept $-\frac{1}{x_1}$. For the line from -4 to $+1$,

$$5y + 2x = 2,$$

intersects the vertical diameter in the point $(0, \frac{2}{5})$. Hence the equation of the line from $\frac{4}{x_1}$ to this point is

$$5y - 2x_1x = 2,$$

and its intersection with the lower tangent gives $-\frac{1}{x_1}$

The projection from A of the intersections of the line from $-\frac{1}{x_1}$ to $\frac{4}{x_1}$ with the circle determines upon the axis of X the two roots of the second quadratic equation, of which, as

already noted, we need only the greater, y_1. This corresponds, as shown by the figure, to the projection of the upper intersection of our transversal with the circle.

Similarly, we obtain the roots of the third quadratic equation. Upon the upper tangent we project from O the intersection of the circle with the straight line which gave upon the axis of X the root $+x_2$. This immediately gives the intercept $\frac{4}{x_2}$, by reason of the correspondence just explained.

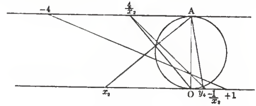


FIG. 7.

If we connect this point with the point where the vertical diameter intersects the line joining -4 above and $+1$ below, we cut off upon the axis of X the segment $-\frac{1}{x_2}$, as desired. If we project that intersection of this transversal with the circle which lies in the positive quadrant from A upon the axis of X, we have constructed the required root y_4 of the third quadratic equation.

We have finally to determine the root z_1 of the fourth quadratic equation and for this purpose to lay off $\frac{4}{y_1}$ above and $\frac{y_4}{y_1}$ below. We solve the first problem in the usual way, by projecting the intersection of the circle with the line connecting A with $+y_1$ below, from O upon the upper tangent, thus obtaining $\frac{4}{y_1}$. For the other segment we connect the point $+4$ above with y_4 below, and then the point thus determined

upon the vertical diameter produced with $\frac{4}{y_1}$. This line cuts off upon the axis of X exactly the segment desired, $\frac{y_4}{y_1}$. For the line a (Fig. 8) has the equation

$$(y_4 - 4)\, y + 2x = 2y_4.$$

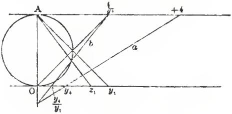


FIG. 8.

It cuts off upon the vertical diameter the segment $\frac{2y_4}{y_4 - 4}$. The equation of the line b is then

$$2y_1x + (y_4 - 4)\, y = 2y_4,$$

and its intersection with the axis of X has the abscissa $\frac{y_4}{y_1}$.

If we project the upper intersection of the line b with the circle from A upon the axis of X, we obtain $z_1 = 2 \cos \frac{2\pi}{17}$. If we desire the simple cosine itself we have only to draw a diameter parallel to the axis of X, on which our last projecting ray cuts off directly $\cos \frac{2\pi}{17}$. A perpendicular erected at this point gives immediately the first and sixteenth vertices of the regular polygon of 17 sides.

The period z_1 was chosen arbitrarily; we might construct in the same way every other period of two terms and so find the remaining cosines. These constructions, made on separate figures so as to be followed more easily, have been combined in a single figure (Fig. 9), which gives the complete construction of the regular polygon of 17 sides.

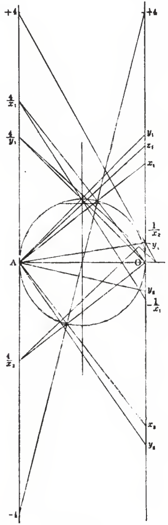

+4
+4
$\frac{4}{x_1}$
$\frac{4}{y_1}$
y_1
z_1
x_1
$\frac{1}{x_2}$
y_1
A
O
y_2
$-\frac{1}{x_1}$
$\frac{4}{x_2}$
x_2
y_2
-4

FIG. 9.

CHAPTER V.

General Considerations on Algebraic Constructions.

1. We shall now lay aside the matter of construction with straight edge and compasses. Before quitting the subject we may mention a new and very simple method of effecting certain constructions, *paper folding.* Hermann Wiener * has shown how by paper folding we may obtain the network of the regular polyhedra. Singularly, about the same time a Hindu mathematician, Sundara Row, of Madras, published a little book, *Geometrical Exercises in Paper Folding* (Madras, Addison & Co., 1893), in which the same idea is considerably developed. The author shows how by paper folding we may construct by points such curves as the ellipse, cissoid, etc.

2. Let us now inquire how to solve geometrically problems whose analytic form is an equation of the third or of higher degree, and in particular, let us see how the ancients succeeded. The most natural method is by means of the conics, of which the ancients made much use. For example, they found that by means of these curves they were enabled to solve the problems of the duplication of the cube and the trisection of the angle. We shall in this place give only a general sketch of the process, making use of the language of modern mathematics for greater simplicity.

Let it be required, for instance, to solve graphically the cubic equation

$$x^3 + ax^2 + bx + c = 0,$$

or the biquadratic,

$$x^4 + ax^3 + bx^2 + cx + d = 0.$$

* See Dyck, *Katalog der Münchener mathematischen Ausstellung von 1893*, Nachtrag, p. 52.

Put $x^2 = y$; our equations become

$$xy + ay + bx + c = 0$$

and
$$y^2 + axy + by + cx + d = 0.$$

The roots of the equations proposed are thus the abscissas of the points of intersection of the two conics.

The equation

$$x^2 = y$$

represents a parabola with axis vertical. The second equation,

$$xy + ay + bx + c = 0,$$

represents an hyperbola whose asymptotes are parallel to the axes of reference (Fig. 10). One of the four points of inter-

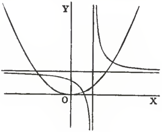


Fig. 10.

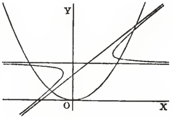


Fig. 11.

section is at infinity upon the axis of Y, the other three at a finite distance, and their abscissas are the roots of the equation of the third degree.

In the second case the parabola is the same. The hyperbola (Fig. 11) has again one asymptote parallel to the axis of X while the other is no longer perpendicular to this axis. The curves now have four points of intersection at a finite distance.

The methods of the ancient mathematicians are given in detail in the elaborate work of M. Cantor, *Geschichte der Mathematik* (Leipzig, 1894, 2d ed.). Especially interesting is Zeuthen, *Die Kegelschnitte im Altertum* (Kopenhagen, 1886, in German edition). As a general compendium we may mention Baltzer, *Analytische Geometrie* (Leipzig, 1882).

3. Beside the conics, the ancients used for the solution of the above-mentioned problems, higher curves constructed for this very purpose. We shall mention here only the *Cissoid* and the *Conchoid*.

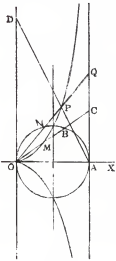


FIG. 12.

The *cissoid of Diocles* (*c.* 150 B.C.) may be constructed as follows (Fig. 12): To a circle draw a tangent (in the figure the vertical tangent on the right) and the diameter perpendicular to it. Draw lines from O, the vertex of the circle thus determined, to points upon the tangent, and lay off from O upon each the segment lying between its intersection with the circle and the tangent. The locus of points so determined is the *cissoid*.

To derive the equation, let r be the radius vector, θ the angle it makes with the axis of X. If we produce r to the tangent on the right, and call the diameter of the circle 1, the total segment equals $\dfrac{1}{\cos\theta}$. The portion cut off by the circle is $\cos\theta$. The difference of the two segments is r, and hence

$$r = \frac{1}{\cos\theta} - \cos\theta = \frac{\sin^2\theta}{\cos\theta}.$$

By transformation of coördinates we obtain the Cartesian equation,

$$(x^2 + y^2)\,x - y^2 = 0.$$

The curve is of the third order, has a cusp at the origin, and is symmetric to the axis of X. The vertical tangent to the circle with which we began our construction is an asymptote. Finally the cissoid cuts the line at infinity in the circular points.

To show how to solve the Delian problem by the use of this curve, we write its equation in the following form:

$$\left(\frac{y}{x}\right)^3 = \frac{y}{1-x}.$$

We now construct the straight line,

$$\frac{y}{x} = \lambda.$$

This cuts off upon the tangent $x = 1$ the segment λ, and intersects the cissoid in a point for which

$$\frac{y}{1-x} = \lambda^3.$$

This is the equation of a straight line passing through the point $y = 0$, $x = 1$, and hence of the line joining this point to the point of the cissoid.

This line cuts off upon the axis of Y the intercept λ^3.

We now see how $\sqrt[3]{2}$ may be constructed. Lay off upon the axis of Y the intercept 2, join this point to the point $x = 1$, $y = 0$, and through its intersection with the cissoid draw a line from the origin to the tangent $x = 1$. The intercept on this tangent equals $\sqrt[3]{2}$.

4. The *conchoid of Nicomedes* (*c.* 150 B.C.) is constructed as follows: Let O be a fixed point, a its distance from a fixed

line. If we pass a pencil of rays through O and lay off on each ray from its intersection with the fixed line in both directions a segment b, the locus of the points so determined is the *conchoid*. According as b is greater or less than a, the origin is a node or a conjugate point; for $b = a$ it is a cusp (Fig. 13).

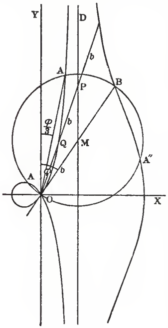


FIG. 13.

Taking for axes of X and Y the perpendicular and parallel through O to the fixed line, we have

$$\frac{r}{x} = \frac{b}{x - a};$$

whence

$$(x^2 + y^2)(x - a)^2 - b^2x^2 = 0.$$

The conchoid is then of the fourth order, has a double point at the origin, and is composed of two branches having for common asymptote the line $x = a$. Further, the factor $(x^2 + y^2)$ shows that the curve passes through the circular points at infinity, a matter of immediate importance.

We may trisect any angle by means of this curve in the following manner: Let $\phi = MOY$ (Fig. 13) be the angle to be divided into three equal parts. On the side OM lay off $OM = b$, an arbitrary length. With M as a center and radius b describe a circle, and through M perpendicular to the axis of X with origin O draw a vertical line representing the asymptote of the conchoid to be constructed. Construct the

conchoid. Connect O with A, the intersection of the circle and the conchoid. Then is $\angle$ AOY one third of $\angle \phi$, as is easily seen from the figure.

Our previous investigations have shown us that the problem of the trisection of the angle is a problem of the third degree. It admits the three solutions

$$\frac{\phi}{3}, \quad \frac{\phi + 2\pi}{3}, \quad \frac{\phi + 4\pi}{3}.$$

Every algebraic construction which solves this problem by the aid of a curve of higher degree must obviously furnish all the solutions. Otherwise the equation of the problem would not be irreducible. These different solutions are shown in the figure. The circle and the conchoid intersect in eight points. Two of them coincide with the origin, two others with the circular points at infinity. None of these can give a solution of the problem. There remain, then, four points of intersection, so that we seem to have one too many. This is due to the fact that among the four points we necessarily find the point B such that $\mathrm{OMB} = 2\,b$, a point which may be determined without the aid of the curve. There actually remain then only three points corresponding to the three roots furnished by the algebraic solution.

5. In all these constructions with the aid of higher algebraic curves, we must consider the practical execution. We need an instrument which shall trace the curve by a continuous movement, for a construction by points is simply a method of approximation. Several instruments of this sort have been constructed; some were known to the ancients. Nicomedes invented a simple device for tracing the conchoid. It is the oldest of the kind besides the straight edge and compasses. (Cantor, I, p. 302.) A list of instruments of more recent construction may be found in Dyck's Katalog, pp. 227–230. 340. and Nachtrag, pp. 42, 43.

PART II.

TRANSCENDENTAL NUMBERS AND THE QUADRATURE OF THE CIRCLE.

CHAPTER I.

Cantor's Demonstration of the Existence of Transcendental Numbers.

1. Let us represent numbers as usual by points upon the axis of abscissas. If we restrict ourselves to rational numbers the corresponding points will fill the axis of abscissas densely throughout (*überall dicht*), *i.e.*, in any interval no matter how small there is an infinite number of such points. Nevertheless, as the ancients had already discovered, the continuum of points upon the axis is not exhausted in this way; between the rational numbers come in the irrational numbers, and the question arises whether there are not distinctions to be made among the irrational numbers.

Let us define first what we mean by *algebraic numbers.* Every root of an algebraic equation

$$a_0\omega^n + a_1\omega^{n-1} + \cdots + a_{n-1}\omega + a_n = 0$$

with integral coefficients is called an algebraic number. Of course we consider only the real roots. Rational numbers occur as a special case in equations of the form

$$a_0\omega + a_1 = 0.$$

We now ask the question: Does the totality of real algebraic numbers form a continuum, or a discrete series such that other numbers may be inserted in the intervals? These new numbers, the so-called *transcendental* numbers, would then be characterized by this property, that they cannot be roots of an algebraic equation with integral coefficients.

This question was answered first by Liouville (*Comptes rendus*, 1844, and Liouville's *Journal*, Vol. XVI, 1851), and in fact the existence of transcendental numbers was demonstrated by him. But his demonstration, which rests upon the theory of continued fractions, is rather complicated. The investigation is notably simplified by using the developments given by Georg Cantor in a memoir of fundamental importance, *Ueber eine Eigenschaft des Inbegriffes reeller algebraischer Zahlen* (Crelle's *Journal*, Vol. LXXVII, 1873). We shall give his demonstration, making use of a more simple notion which Cantor, under a different form, it is true, suggested at the meeting of naturalists in Halle, 1891.

2. The demonstration rests upon the fact that algebraic numbers form a *countable* mass, while transcendental numbers do not. By this Cantor means that the former can be arranged in a certain order so that each of them occupies a definite place, is numbered, so to speak. This proposition may be stated as follows:

The manifoldness of real algebraic numbers and the manifoldness of positive integers can be brought into a one-to-one correspondence.

We seem here to meet a contradiction. The positive integers form only a portion of the algebraic numbers; since each number of the first can be associated with one and one only of the second, the part would be equal to the whole. This objection rests upon a false analogy. The proposition that the part is always less than the whole is not true for

infinite masses. It is evident, for example, that we may establish a one-to-one correspondence between the aggregate of positive integers and the aggregate of positive even numbers, thus:

$$\begin{array}{llllll} 0 & 1 & 2 & 3 \cdots & n \cdots \\ 0 & 2 & 4 & 6 \cdots & 2n \cdots \end{array}$$

In dealing with infinite masses, the words *great* and *small* are inappropriate. As a substitute, Cantor has introduced the word *power* (*Mächtigkeit*), and says: *Two infinite masses have the same power when they can be brought into a one-to-one correspondence with each other.* The theorem which we have to prove then takes the following form: *The aggregrate of real algebraic numbers has the same power as the aggregate of positive integers.*

We obtain the aggregate of real algebraic numbers by seeking the real roots of all algebraic equations of the form

$$a_0\omega^n + a_1\omega^{n-1} + \cdots + a_{n-1}\omega + a_n = 0;$$

all the a's are supposed prime to one another, a_0 positive, and the equation irreducible. To arrange the numbers thus obtained in a definite order, we consider their *height* N as defined by

$$N = n - 1 + |a_0| + |a_1| + \cdots + |a_n|,$$

$|a_i|$ representing the absolute value of a_i, as usual. To a given number N corresponds a finite number of algebraic equations. For, N being given, the number n has certainly an upper limit, since N is equal to $n-1$ increased by positive numbers; moreover, the difference $N-(n-1)$ is a sum of positive numbers prime to one another, whose number is obviously finite.

N	n	$\lvert a_0\rvert$	$\lvert a_1\rvert$	$\lvert a_2\rvert$	$\lvert a_3\rvert$	$\lvert a_4\rvert$	EQUATION.	$\phi(N)$	ROOTS.
1	1	1	0				$x = 0$	1	0
	2	0	0	0			—		
2	1	2	0				—	2	-1
		1	1				$x \pm 1 = 0$		$+1$
	2	1	0	0			—		
3	1	3	0				—	4	-2
		2	1				$2x \pm 1 = 0$		$-\frac{1}{2}$
		1	2				$x \pm 2 = 0$		$+\frac{1}{2}$
	2	2	0	0			—		$+2$
		1	1	0			—		
		1	0	1			—		
	3	1	0	0	0		—		
4	1	4	0				—	12	-3
		3	1				$3x \pm 1 = 0$		-1.61803
		2	2				—		-1.41421
		1	3				$x \pm 3 = 0$		-0.70711
	2	3	0	0			—		-0.61803
		2	1	0			—		-0.33333
		2	0	1			$2x^2 - 1 = 0$		$+0.33333$
		1	2	0			—		$+0.61803$
		1	1	1			$x^2 \pm x - 1 = 0$		$+0.70711$
		1	0	2			$x^2 - 2 = 0$		$+1.41421$
	3	2	0	0	0		—		$+1.61803$
		1	1	0	0		—		$+3$
		1	0	1	0		—		
		1	0	0	1		—		
	4	1	0	0	0	0	—		

Among these equations we must discard those that are reducible, which presents no theoretical difficulty. Since the number of equations corresponding to a given value of

N is limited, there corresponds to a determinate N only a finite mass of algebraic numbers. We shall designate this by $\phi(N)$. The table contains the values of $\phi(1)$, $\phi(2)$, $\phi(3)$, $\phi(4)$, and the corresponding algebraic numbers ω.

We arrange now the algebraic numbers according to their height, N, and the numbers corresponding to a single value of N in increasing magnitude. We thus obtain all the algebraic numbers, each in a determinate place. This is done in the last column of the accompanying table. It is, therefore, evident that algebraic numbers can be counted.

3. We now state the general proposition:

In any portion of the axis of abscissas, however small, there is an infinite number of points which certainly do not belong to a given countable mass.

Or, in other words:

The continuum of numerical values represented by a portion of the axis of abscissas, however small, has a greater power than any given countable mass.

This amounts to affirming the existence of transcendental numbers. It is sufficient to take as the countable mass the aggregate of algebraic numbers.

To demonstrate this theorem we prepare a table of algebraic numbers as before and write in it all the numbers in the form of decimal fractions. None of these will end in an infinite series of 9's. For the equality

$$1 = 0.999 \cdots 9 \cdots$$

shows that such a number is an exact decimal. If now we can construct a decimal fraction which is not found in our table and does not end in an infinite series of 9's it will certainly be a transcendental number. By means of a very simple process indicated by Georg Cantor we can find not only one but infinitely many transcendental numbers, even

when the domain in which the number is to lie is very small Suppose, for example, that the first five decimals of the number are given. Cantor's process is as follows.

Take for 6th decimal a number different from 9 and from the 6th decimal of the *first algebraic* number, for 7th decimal a number different from 9 and from the 7th decimal of the *second algebraic* number, etc. In this way we obtain a decimal fraction which will not end in an infinite series of 9's and is certainly not contained in our table. The proposition is then demonstrated.

We see by this that (if the expression is allowable) there are far more transcendental numbers than algebraic. For when we determine the unknown decimals, avoiding the 9's, we have a choice among eight different numbers; we can thus form, so to speak, 8^∞ transcendental numbers, even when the domain in which they are to lie is as small as we please.

CHAPTER II.

Historical Survey of the Attempts at the Computation and Construction of π.

In the next chapter we shall prove that the number π belongs to the class of transcendental numbers whose existence was shown in the preceding chapter. The proof was first given by Lindemann in 1882, and thus a problem was definitely settled which, so far as our knowledge goes, has occupied the attention of mathematicians for nearly 4000 years, the problem of the quadrature of the circle.

For, if the number π is not algebraic, it certainly cannot be constructed by means of straight edge and compasses. *The quadrature of the circle in the sense understood by the ancients is then impossible.* It is extremely interesting to follow the fortunes of this problem in the various epochs of science, as ever new attempts were made to find a solution with straight edge and compasses, and to see how these necessarily fruitless efforts worked for advancement in the manifold realm of mathematics.

The following brief historical survey is based upon the excellent work of Rudio: *Archimedes, Huygens, Lambert, Legendre, Vier Abhandlungen über die Kreismessung*, Leipzig, 1892. This book contains a German translation of the investigations of the authors named. While the mode of presentation does not touch upon the modern methods here discussed, the book includes many interesting details which are of practical value in elementary teaching.

1. Among the attempts to determine the ratio of the diameter to the circumference we may first distinguish the *empirical stage*, in which the desired end was to be attained by measurement or by direct estimation.

One of the oldest known mathematical documents, the Rhind Papyrus (*c.* 1650 B.C.), contains the problem in the well-known form, to transform a circle into a square of equal area. The writer of the papyrus lays down the following rule: Cut off $\frac{1}{9}$ of a diameter and construct a square upon the remainder; this has the same area as the circle. The value of π thus obtained is $(\frac{16}{9})^2 = 3.16 \cdots$, not very inaccurate. Much less accurate is the value $\pi = 3$, used in the Bible (1 Kings, 7. 23, 2 Chronicles, 4. 2).

2. The Greeks rose above this empirical standpoint, and especially Archimedes, who, in his work κύκλου μέτρησις, computed the area of the circle by the aid of inscribed and circumscribed polygons, as is still done in the schools. His method remained in use till the invention of the differential calculus; it was especially developed and rendered practical by Huygens (d. 1654) in his work, *De circuli magnitudine inventa.*

As in the case of the duplication of the cube and the trisection of the angle the Greeks sought also to effect the quadrature of the circle by the help of higher curves.

Consider for example the curve $y = \sin^{-1} x$, which represents the sinusoid with axis vertical. Geometrically, π appears as a particular ordinate of this curve; from the standpoint of the theory of functions, as a particular value of our transcendental function. Any apparatus which describes a transcendental curve we shall call a transcendental apparatus. A transcendental apparatus which traces the sinusoid gives us a geometric construction of π.

In modern language the curve $y = \sin^{-1} x$ is called an

integral curve because it can be defined by means of the integral of an algebraic function,

$$y = \int_0 \frac{dx}{\sqrt{1-x^2}}.$$

The ancients called such a curve a *quadratrix* or *τετραγωνίζουσα*. The best known is the *quadratrix of Dinostratus* (*c.* 350 B.C.) which, however, had already been constructed by Hippias of Elis (*c.* 420 B.C.) for the trisection of an angle. Geometrically it may be defined as follows. Having given a circle and two perpendicular radii OA and OB, two points M and L move with constant velocity, one upon the radius OB, the other upon the arc AB (Fig. 14). Starting at the same time at O and A, they arrive simultaneously at B. The point of intersection P of OL and the parallel to OA through M describes the quadratrix.

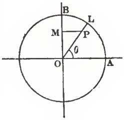


FIG. 14.

From this definition it follows that y is proportional to θ. Further, since for $y = 1$, $\theta = \frac{\pi}{2}$ we have

$$\theta = \frac{\pi}{2} y;$$

and from $\theta = \tan^{-1}\frac{y}{x}$ the equation of the curve becomes

$$\frac{y}{x} = \tan \frac{\pi}{2} y.$$

It meets the axis of X at the point whose abscissa is

$$x = \lim \frac{y}{\tan \frac{\pi}{2} y}, \text{ for } y = 0;$$

hence $$x = \frac{2}{\pi}.$$

According to this formula the radius of the circle is the mean proportional between the length of the quadrant and the abscissa of the intersection of the quadratrix with the axis of X. This curve can therefore be used for the rectification and hence also for the quadrature of the circle. This use of the quadratrix amounts, however, simply to a geometric formulation of the problem of rectification so long as we have no apparatus for describing the curve by continuous movement.

Fig. 15 gives an idea of the form of the curve with the branches obtained by taking values of θ greater than π or

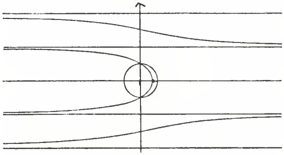


FIG. 15.

less than $-\pi$. Evidently the quadratrix of Dinostratus is not so convenient as the curve $y = \sin^{-1} x$, but it does not appear that the latter was used by the ancients.

3. The period from 1670 to 1770, characterized by the names of Leibnitz, Newton, and Euler, saw the rise of modern analysis. Great discoveries followed one another in such an almost unbroken series that, as was natural, critical rigor fell into the background. For our purposes the development

of the theory of series is especially important. Numerous methods were deduced for approximating the value of π. It will suffice to mention the so-called *Leibnitz series* (known, however, before Leibnitz):

$$\frac{\pi}{4} = 1 - \tfrac{1}{3} + \tfrac{1}{5} - \tfrac{1}{7} \cdots$$

This same period brings the discovery of the mutual dependence of e and π. The number e, natural logarithms, and hence the exponential function, are first found in principle in the works of Napier (1614). This number seemed at first to have no relation whatever to the circular functions and the number π until Euler had the courage to make use of imaginary exponents. In this way he arrived at the celebrated formula

$$e^{ix} = \cos x + i \sin x,$$

which, for $x = \pi$, becomes

$$e^{i\pi} = -1.$$

This formula is certainly one of the most remarkable in all mathematics. The modern proofs of the transcendence of π are all based on it, since the first step is always to show the transcendence of e.

4. After 1770 critical rigor gradually began to resume its rightful place. In this year appeared the work of Lambert: *Vorläufige Kenntnisse für die so die Quadratur des Cirkuls suchen.* Among other matters the irrationality of π is discussed. In 1794 Legendre, in his *Éléments de géométrie*, showed conclusively that π and π^2 are irrational numbers.

5. But a whole century elapsed before the question was investigated from the modern point of view. The starting-point was the work of Hermite: *Sur la fonction exponentielle* (*Comptes rendus*, 1873, published separately in 1874). The transcendence of e is here proved.

An analogous proof for π, closely related to that of Hermite, was given by Lindemann: *Ueber die Zahl* π (*Mathematische Annalen*, XX, 1882. See also the Proceedings of the Berlin and Paris academies).

The question was then settled for the first time, but the investigations of Hermite and Lindemann were still very complicated.

The first simplification was given by Weierstrass in the *Berliner Berichte* of 1885. The works previously mentioned were embodied by Bachmann in his text-book, *Vorlesungen über die Natur der Irrationalzahlen*, 1892.

But the spring of 1893 brought new and very important simplifications. In the first rank should be named the memoirs of Hilbert in the *Göttinger Nachrichten.* Still Hilbert's proof is not absolutely elementary: there remain traces of Hermite's reasoning in the use of the integral

$$\int_0^\infty z^\rho e^{-z} dz = \rho\,!$$

But Hurwitz and Gordan soon showed that this transcendental formula could be done away with (*Göttinger Nachrichten; Comptes rendus;* all three papers are reproduced with some extensions in *Mathematische Annalen*, Vol. XLIII).

The demonstration has now taken a form so elementary that it seems generally available. In substance we shall follow Gordan's mode of treatment.

CHAPTER III.

The Transcendence of the Number e.

1. We take as the starting-point for our investigation the well-known series

$$e^x = 1 + \frac{x}{1} + \frac{x^2}{2!} + \cdots \frac{x^n}{n!} + \cdots,$$

which is convergent for all finite values of x. The difference between practical and theoretical convergence should here be insisted on. Thus, for $x = 1000$ the calculation of e^{1000} by means of this series would obviously not be feasible. Still the series certainly converges theoretically; for we easily see that after the 1000th term the factorial $n!$ in the denominator increases more rapidly than the power which occurs in the numerator. This circumstance that $\frac{x^n}{n!}$ has for any finite value of x the limit zero when n becomes infinite has an important bearing upon our later demonstrations.

We now propose to establish the following proposition:

The number e *is not an algebraic number, i.e.*, an equation with integral coefficients of the form

$$F(e) = C_0 + C_1 e + C_2 e^2 + \cdots + C_n e^n = 0$$

is impossible. The coefficients C_i may be supposed prime to one another.

We shall use the indirect method of demonstration, showing that the assumption of the above equation leads to an absurdity. The absurdity may be shown in the following

way. We multiply the members of the equation $F(e)=0$ by a certain integer M so that

$$MF(e) = MC_0 + MC_1e + MC_2e^2 + \cdots + MC_ne^n = 0.$$

We shall show that the number M can be chosen so that

(1) Each of the products Me, Me^2, $\cdots Me^n$ may be separated into an entire part M_κ and a fractional part ϵ_κ, and our equation takes the form

$$\begin{aligned} MF(e) = MC_0 + M_1C_1 + M_2C_2 + \cdots + M_nC_n \\ + C_1\epsilon_1 + C_2\epsilon_2 + \cdots + C_n\epsilon_n = 0; \end{aligned}$$

(2) The integral part

$$MC_0 + M_1C_1 + \cdots + M_nC_n$$

is not zero. This will result from the fact that when divided by a prime number it gives a remainder different from zero;

(3) The expression

$$C_1\epsilon_1 + C_2\epsilon_2 + \cdots + C_n\epsilon_n$$

can be made as small a fraction as we please.

These conditions being fulfilled, the equation assumed is manifestly impossible, since the sum of an integer different from zero, and a proper fraction, cannot equal zero.

The salient point of the proof may be stated, though not quite accurately, as follows:

With an exceedingly small error we may assume e, e^2, $\cdots e^n$ proportional to integers which certainly do not satisfy our assumed equation.

2. We shall make use in our proof of a symbol h^r and a certain polynomial $\phi(x)$.

The symbol h^r is simply another notation for the factorial $r!$ Thus, we shall write the series for e^x in the form

$$e^x = 1 + \frac{x}{h} + \frac{x^2}{h^2} + \cdots + \frac{x^n}{h^n} + \cdots$$

The symbol has no deeper meaning; it simply enables us to write in more compact form every formula containing powers and factorials.

Suppose, *e.g.*, we have given a developed polynomial

$$f(x) = \sum_r c_r x^r.$$

We represent by $f(h)$, and write under the form $\sum_r c_r h^r$, the sum

$$c_1 \cdot 1 + c_2 \cdot 2! + c_3 \cdot 3! + \cdots + c_n \cdot n!$$

But if $f(x)$ is not developed, then to calculate $f(h)$ is to develop this polynomial in powers of h and finally replace h^r by $r!$. Thus, for example,

$$f(k+h) = \sum_r c_r (k+h)^r = \sum_r c'_r \cdot h^r = \sum_r c'_r \cdot r!,$$

the c'_r depending on k.

The polynomial $\phi(x)$ which we need for our proof is the following remarkable expression

$$\phi(x) = x^{p-1} \frac{[(1-x)(2-x)\cdots(n-x)]^p}{(p-1)!},$$

where p is a prime number, n the degree of the algebraic equation assumed to be satisfied by e. We shall suppose p greater than n and $|C_0|$, and later we shall make it increase without limit.

To get a geometric picture of this polynomial $\phi(x)$ we construct the curve

$$y = \phi(x).$$

At the points $x = 1, 2, \cdots n$ the curve has the axis of X as an inflexional tangent, since it meets it in an odd number of points, while at the origin the axis of X is tangent without inflexion. For values of x between 0 and n the curve remains in the neighborhood of the axis of X; for greater values of x it recedes indefinitely.

Of the function $\phi(x)$ we will now establish three important properties :

1. x *being supposed given and* p *increasing without limit,* $\phi(x)$ *tends toward zero, as does also the sum of the absolute values of its terms.*

Put $u = x(1-x)(2-x)\cdots(n-x)$; we may then write

$$\phi(x)=\frac{u^{p-1}}{(p-1)!}\,\frac{u}{x},$$

which for p infinite tends toward zero.

To have the sum of the absolute values of $\phi(x)$ it is sufficient to replace $-x$ by $|x|$ in the undeveloped form of $\phi(x)$. The second part is then demonstrated like the first.

2. h *being an integer,* $\phi(h)$ *is an integer not divisible by* p *and therefore different from zero.*

Develop $\phi(x)$ in increasing powers of x, noticing that the terms of lowest and highest degree respectively are of degree $p-1$ and $np+p-1$. We have

$$\phi(x)=\sum_{r=p-1}^{r=np+p-1} c_r x^r=\frac{c' x^{p-1}}{(p-1)!}+\frac{c'' x^{p}}{(p-1)!}+\cdots\pm\frac{x^{np+p-1}}{(p-1)!}.$$

Hence

$$\phi(h)=\sum_{r=p-1}^{r=np+p-1} c_r h^r.$$

Leaving out of account the denominator $(p-1)!$, which occurs in all the terms, the coefficients c_r are integers. This denominator disappears as soon as we replace h^r by $r!$, since the factorial of least degree is $h^{p-1}=(p-1)!$. All the terms of the development after the first will contain the factor p. As to the first, it may be written

$$\frac{(1\cdot2\cdot3\cdots n)^p\cdot(p-1)!}{(p-1)!}=(n!)^p$$

and is certainly not divisible by p since $p>n$.

Therefore $\qquad \phi(h)\equiv(n!)^p \pmod{p}$,

and hence $\qquad \phi(h)\not\equiv 0.$

Moreover, $\phi(h)$ is a very large number; even its last term alone is very large, viz.:

$$\frac{(np+p-1)!}{(p-1)!} = p(p+1)\cdots(np+p-1).$$

3. h *being an integer, and* k *one of the numbers* $1, 2 \cdots n$, $\phi(h+k)$ *is an integer divisible by* p.

We have $\quad \phi(h+k) = \sum_r c_r (h+k)^r = \sum_r c'_r h^r,$

a formula in which we are to replace h^r by $r!$ only after having arranged the development in increasing powers of h.

According to the rules of the symbolic calculus, we have first

$$\phi(h+k) \\ = (h+k)^{p-1} \frac{[(1-k-h)(2-k-h)\cdots(-h)\cdots(n-k-h)]^p}{(p-1)!}.$$

One of the factors in the brackets reduces to $-h$; hence the term of lowest degree in h in the development is of degree p. We may then write

$$\phi(h+k) = \sum_{r=p}^{r=np+p-1} c'_r h^r.$$

The coefficients still have for numerators integers and for denominator $(p-1)!$. As already explained, this denominator disappears when we replace h^r by $r!$. But now all the terms of the development are divisible by p; for the first may be written

$$\frac{(-1)^{kp}\cdot k^{p-1}[(k-1)!(n-k)!]^p\cdot p!}{(p-1)!} \\ = (-1)^{kp} k^{p-1}[(k-1)!\cdot(n-k)!]^p\cdot p.$$

$\phi(h+k)$ is then divisible by p.

3. We can now show that the equation

$$F(e) = C_0 + C_1 e + C_2 e^2 + \cdots + C_n e^n = 0$$

is impossible.

For the number M, by which we multiply the members of this equation, we select $\phi(h)$, so that

$$\phi(h)F(e) = C_0\phi(h) + C_1\phi(h)e + C_2\phi(h)e^2 + \cdots + C_n\phi(h)e^n.$$

Let us try to decompose any term, such as $C_k\phi(h)e^k$, into an integer and a fraction. We have

$$e^k \cdot \phi(h) = e^k \sum_r c_r h^r.$$

Considering the series development of e^k, any term of this sum, omitting the constant coefficient, has the form

$$e^k \cdot h^r = h^r + \frac{h^r \cdot k}{1} + \frac{h^r \cdot k^2}{2!} + \cdots + \frac{h^r \cdot k^r}{r!} + \frac{h^r \cdot k^{r+1}}{(r+1)!} + \cdots$$

Replacing h^r by $r!$, or what amounts to the same thing, by one of the quantities

$$rh^{r-1},\ r(r-1)h^{r-2}\cdots,\ r(r-1)\cdots 3\cdot h^2,\ r(r-1)\cdots 2\cdot h,$$

and simplifying the successive fractions,

$$e^k \cdot h^r = h^r + \frac{r}{1}\cdot h^{r-1}k + \frac{r(r-1)}{2!}h^{r-2}k^2 + \cdots + \frac{r}{1}hk^{r-1} + k^r$$
$$+ k^r\left[\frac{k}{r+1} + \frac{k^2}{(r+1)(r+2)} + \cdots\right].$$

The first line has the same form as the development of $(h+k)^r$; in the parenthesis of the second line we have the series

$$0 + \frac{k}{r+1} + \frac{k^2}{(r+1)(r+2)} + \cdots$$

whose terms are respectively less than those of the series

$$e^k = 1 + k + \frac{k^2}{2!} + \frac{k^3}{3!} + \cdots$$

The second line in the expansion of $e^k \cdot h^r$ may therefore be represented by an expression of the form

$$q_{r,k} \cdot e^k \cdot k^r,$$

$q_{r,k}$ being a proper fraction.

Effecting the same decomposition for each term of the sum

$$e^k \sum c_r h^r$$

it takes the form

$$e^k \sum_r c_r h^r = \sum_r c_r (h+k)^r + e^k \sum_r q_{r,k} c_r k^r.$$

The first part of this sum is simply $\phi(h+k)$; this is a number divisible by p (**2**, 3). Further (**2**, 1),

$$\phi(k) = \sum_r |c_r k^r|$$

tends toward zero when p becomes infinite: the same is true *a fortiori* of $\sum_r q_{r,k} c_r k^r$, and also, since e^k is a finite quantity, of $e^k \sum_r q_{r,k} c_r k^r$, which we may represent by ϵ_k.

The term under consideration, $C_k e^k \phi(h)$, has then been put under the form of an integer $C_k \phi(h+k)$ and a quantity $C_k \epsilon_k$ which, by a suitable choice of p, may be made as small as we please.

Proceeding similarly with all the terms, we get finally

$$\begin{aligned} F(e)\,\phi(h) = C_0 \phi(h) + C_1 \phi(h+1) + \cdots + C_n \phi(h+n) \\ + C_1 \epsilon_1 + C_2 \epsilon_2 + \cdots + C_n \epsilon_n. \end{aligned}$$

It is now easy to complete the demonstration. All the terms of the first line after the first are divisible by p; for the first, $|C_0|$ is less than p; $\phi(h)$ is not divisible by p; hence $C_0\phi(h)$ is not divisible by the prime number p. Consequently the sum of the numbers of the first line is not zero.

The numbers of the second line are finite in number; each of them can be made smaller than any given number by a suitable choice of p; and therefore the same is true of their sum.

Since an integer not zero and a fraction cannot have zero for a sum, the assumed equation is impossible.

Thus, the transcendence of e, or Hermite's Theorem, is demonstrated.

CHAPTER IV.

The Transcendence of the Number π.

1. The demonstration of the transcendence of the number π given by Lindemann is an extension of Hermite's proof in the case of e. While Hermite shows that an integral equation of the form

$$C_0 + C_1 e + C_2 e^2 + \cdots + C_n e^n = 0$$

cannot exist, Lindemann generalizes this by introducing in place of the powers e, $e^2 \cdots$ sums of the form

$$\begin{gathered} e^{k_1} + e^{k_2} + \cdots + e^{k_N} \\ e^{l_1} + e^{l_2} + \cdots + e^{l_{N'}} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \end{gathered}$$

where the k's are associated algebraic numbers, *i.e.*, roots of an algebraic equation, with integral coefficients, of the degree N; the l's roots of an equation of degree N', etc. Moreover, some or all of these roots may be imaginary.

Lindemann's general theorem may be stated as follows:

The number e *cannot satisfy an equation of the form*

$$\begin{aligned} (1) \quad C_0 + C_1(e^{k_1} + e^{k_2} + \cdots + e^{k_N}) \\ + C_2(e^{l_1} + e^{l_2} + \cdots + e^{l_{N'}}) + \cdots = 0 \end{aligned}$$

where the coefficients C_i *are integers and the exponents* $k_1, l_1, \cdots$ *are respectively associated algebraic numbers.*

The theorem may also be stated:

The number e *is not only not an algebraic number and therefore a transcendental number simply, but it is also not an interscendental* number and therefore a transcendental number of higher order.*

* Leibnitz calls a function x^λ, where λ is an algebraic irrational, an interscendental function.

Let

$$ax^{\nu} + a_1x^{\nu-1} + \cdots + a_{\nu} = 0$$

be the equation having for roots the exponents k_i;

$$bx^{\nu'} + b_1x^{\nu'-1} + \cdots + b_{\nu'} = 0$$

that having for roots the exponents l_i, etc. These equations are not necessarily irreducible, nor the coefficients of the first terms equal to 1. It follows that the symmetric functions of the roots which alone occur in our later developments need not be integers.

In order to obtain integral numbers it will be sufficient to consider symmetric functions of the quantities

$$ak_1,\ ak_2,\ \cdots ak_{\nu},$$
$$bl_1,\ bl_2,\ \cdots bl_{\nu'},\ \text{etc.}$$

These numbers are roots of the equations

$$y^{\nu} + a_1y^{\nu-1} + a_2ay^{\nu-2} + \cdots + a_{\nu}a^{\nu-1} = 0,$$
$$y^{\nu'} + b_1y^{\nu'-1} + b_2by^{\nu'-2} + \cdots + b_{\nu'}b^{\nu'-1} = 0,\ \text{etc.}$$

These quantities are integral associated algebraic numbers, and their rational symmetric functions real integers.

We shall now follow the same course as in the demonstration of Hermite's theorem.

We assume equation (1) to be true; we multiply both members by an integer M; and we decompose each sum, such as

$$M(e^{k_1} + e^{k_2} + \cdots + e^{k_{\nu}}),$$

into an integral part and a fraction, thus

$$M(e^{k_1} + e^{k_2} + \cdots + e^{k_{\nu}}) = M_1 + \epsilon_1,$$
$$M(e^{l_1} + e^{l_2} + \cdots + e^{l_{\nu'}}) = M_2 + \epsilon_2,$$
$$\cdots\cdots\cdots\cdots$$

Our equation then becomes

$$C_0M + C_1M_1 + C_2M_2 + \cdots$$
$$+ C_1\epsilon_1 + C_2\epsilon_2 + \cdots = 0.$$

We shall show that with a suitable choice of M the sum of the quantities in the first line represents an integer not divisible by a certain prime number p, and consequently different from zero; that the fractional part can be made as small as we please, and thus we come upon the same contradictiou as before.

2. We shall again use the symbol $h^r = r!$ and select as the multiplier the quantity $M = \psi(h)$, where $\psi(x)$ is a generalization of $\phi(x)$ used in the preceding chapter, formed as follows:

$$\psi(x) = \frac{x^{p-1}}{(p-1)!}[(k_1 - x)(k_2 - x)\cdots(k_N - x)]^p \cdot a^{Np} \cdot a^{N'p} \cdot a^{N''p} \cdots$$
$$\cdot [(l_1 - x)(l_2 - x)\cdots(l_{N'} - x)]^p \cdot b^{Np} \cdot b^{N'p} \cdot b^{N''p} \cdots$$
$$\cdots\cdots\cdots$$
$$\cdots\cdots\cdots$$

where p is a prime number greater than the absolute value of each of the numbers

$$C_0,\ a,\ b,\ \cdots,\ a_N,\ b_{N'},\ \cdots$$

and later will be assumed to increase without limit. As to the factors a^{Np}, $b^{N'p}$, $\cdots$, they have been introduced so as to have in the development of $\psi(x)$ symmetric functions of the quantities

$$ak_1,\ ak_2,\ \cdots,\ ak_N,$$
$$bl_1,\ bl_2,\ \cdots,\ bl_{N'},$$
$$\cdots\cdots$$

that is, rational integral numbers. Later on we shall have to develop the expressions

$$\sum_\nu \psi(k_\nu + h), \quad \sum_\nu \psi(l_\nu + h), \ \cdots$$

The presence of these same factors will still be necessary if we wish the coefficients of these developments to be integers each divided by $(p-1)!$.

1. $\psi(h)$ *is an integral number, not divisible by* p *and consequently different from zero.*

Arranging $\psi(h)$ in increasing powers of h, it takes the form

$$\psi(h) = \sum_{r=p-1}^{r=\kappa p+\kappa' p+\cdots+p-1} c_r h^r.$$

In this development all the coefficients have integral numerators and the common denominator $(p-1)!$.

The coefficient of the first term h^{p-1} may be written

$$\frac{1}{(p-1)!}(ak_1\cdot ak_2\cdot\cdot\cdot ak_\nu)^p a^{\kappa' p} a^{\kappa'' p}\cdots$$
$$(bl_1\cdot bl_2\cdot\cdot\cdot bl_{\kappa'})^p b^{\kappa p} b^{\kappa'' p}\cdots$$
$$\cdots\cdots\cdots$$
$$=\frac{1}{(p-1)!}(-1)^{\kappa p+\kappa' p+\cdots}(a_\kappa a^{\kappa-1})^p a^{\kappa' p} a^{\kappa'' p}\cdots(b_{\kappa'} b^{\kappa'-1})^p b^{\kappa p} b^{\kappa'' p}\cdots$$

If in this term we replace h^{p-1} by its value $(p-1)!$ the denominator disappears. According to the hypotheses made regarding the prime number p, no factor of the product is divisible by p and hence the product is not.

The second term $c_p h^p$ becomes likewise an integer when we replace h^p by $p!$ but the factor p remains, and so for all of the following terms. Hence $\psi(h)$ is an integer not divisible by p.

2. *For* x, *a given finite quantity, and* p *increasing without limit,* $\psi(x)=\sum_r c_r x^r$ *tends toward zero, as does also the sum* $\sum_r |c_r x^r|$.

We may write

$$\psi(x) = \sum c_r x^r$$
$$= \frac{x^{p-1}}{(p-1)!}\left[a^\kappa a^\kappa\cdots b^\kappa b^\kappa (k_1-x)(k_2-x)\cdots(k_\kappa-x)\right.$$
$$\left.(l_1-x)(l_2-x)\cdots(l_{\kappa'}-x)\cdots\right]^p.$$

Since for x of given value the expression in brackets is a constant, we may replace it by K. We then have

$$\psi(x) = \frac{(xK)^{p-1}}{(p-1)!}K,$$

a quantity which tends toward zero as p increases indefinitely.

The same reasoning will apply when each term of $\psi(x)$ is replaced by its absolute value.

3. *The expression* $\sum_{\nu=1}^{\nu=N} \psi(k_\nu + h)$ *is an integer divisible by* p.

We have

$$\psi(k_\nu + h) = \frac{a^p (k_\nu + h)^{p-1}}{(p-1)!} b^{Np} b^{N''p} \cdots$$
$$\cdot\, a^{(N-1)p} [(k_1 - k_\nu - h)(k_2 - k_\nu - h) \cdots (-h) \cdots (k_N - k_\nu - h)]^p$$
$$\cdot\, a^{N'p} b^{N'p} [(l_1 - k_\nu - h)(l_2 - k_\nu - h) \cdots (l_{N'} - k_\nu - h)]^p$$
$$\cdots\cdots\cdots\cdots$$

The νth factor of the expression in brackets in the second line is $-h$, and hence the term of lowest degree in h is h^p.

Consequently

$$\psi(k_\nu + h) = \sum_{r=p}^{r=Np+N'p+\cdots+p-1} c'_r h^r,$$

whence

$$\sum_{\nu=1}^{\nu=N} \psi(k_\nu + h) = \sum_{r=p}^{r=Np+N p+\cdots+p-1} C'_r h^r.$$

The numerators of the coefficients C'_r are rational and integral, for they are integral symmetric functions of the quantities

$$\begin{array}{cccc} ak_1, & ak_2, & \cdots, & ak_N, \\ bl_1, & bl_2, & \cdots, & bl_N, \\ \cdots & \cdots & \cdots & \cdots \end{array}$$

and their common denominator is $(p-1)!$.

If we replace h^r by $r!$ the denominator disappears from all the coefficients, the factor p remains in every term, and hence the sum is an integer divisible by p.

Similarly for

$$\sum_{\nu=1}^{\nu=N} \psi(l_\nu + h) \cdots$$

We have thus established three properties of $\psi(x)$ analogous to those demonstrated for $\phi(x)$ in connection with Hermite's theorem.

3. We now return to our demonstration that the assumed equation

(1) $C_0+C_1(e^{k_1}+e^{k_2}+\cdots+e^{k_N})+C_2(e^{l_1}+e^{l_2}+\cdots e^{l_{N'}})+\cdots=0$

cannot be true. For this purpose we multiply both members by $\psi(h)$, thus obtaining

$$C_0\psi(h)+C_1[e^{k_1}\psi(h)+e^{k_2}\psi(h)+\cdots+e^{k_N}\psi(h)]+\cdots=0,$$

and try to decompose each of the expressions in brackets into a whole number and a fraction. The operation will be a little longer than before, for k may be a complex number of the form $k=k'+ik''$. We shall need to introduce $|k|=+\sqrt{k'^2+k''^2}$.

One term of the above sum is

$$e^k\cdot\psi(h)=e^k\sum_r c_r h^r=\sum_r c_r\cdot e^k\cdot h^r.$$

The product $e^k\cdot h^r$ may be written, as shown before,

$$e^k\cdot h^r=(h+k)^r+k^r\left[\frac{k}{r+1}+\frac{k^2}{(r+1)(r+2)}+\cdots\right]$$

The absolute value of every term of the series.

$$0+\frac{k}{r+1}+\frac{k^2}{(r+1)(r+2)}+\cdots$$

is less than the absolute value of the corresponding term in the series

$$e^k=1+\frac{k}{1}+\frac{k^2}{2!}+\cdots$$

Hence
$$\left|\frac{k}{r+1}+\frac{k^2}{(r+1)(r+2)}+\cdots\right|<e^{|k|}$$

or
$$\frac{k}{r+1}+\frac{k^2}{(r+1)(r+2)}+\cdots=q_{r,k}e^{|k|},$$

$q_{r,k}$ being a complex quantity whose absolute value is less than 1.

We may then write

$$\begin{aligned}e^k\cdot\psi(h)&=\sum_r c_r e^k h^r=\sum_r c_r(h+k)^r+\sum_r c_r q_{r,k}k^r e^{|k|}\\&=\psi(h+k)+\sum_r c_r q_{r,k}k^r\cdot e^{|k|}\end{aligned}$$

By giving k in succession the indices $1, 2, \cdots N$, and forming the sum the equation becomes

$$e^{k_1}\psi(h) + e^{k_2}\psi(h) + \cdots + e^{k_N}\psi(h)$$
$$= \sum_{\nu=1}^{\nu=N} \psi(k_\nu + h) + \sum_{\nu=1}^{\nu=N} \{ e^{|k_\nu|} \sum_r c_r k^r_\nu q_{r,k_\nu} \}.$$

Proceeding similarly with all the other sums, our equation takes the form

$$(2)\quad C_0\psi(h) + C_1 \sum_{\nu=1}^{\nu=N} \psi(k_\nu + h) + C_2 \sum_{\nu=1}^{\nu=N'} \psi(l_\nu + h) + \cdots$$
$$+ C_1 \sum_{\nu=1}^{\nu=N} \sum_r e^{|k_\nu|} c_r k^r_\nu q_{r,k_\nu} + C_2 \sum_{\nu=1}^{\nu=N'} e^{|l_\nu|} c'_r l^r_\nu q_{r\, l_\nu} + \cdots = 0.$$

By 2, 2 we can make $\sum_r |c_r k^r|$ as small as we please by taking p sufficiently great. Since $|q_{r,k}| < 1$, this will be true *a fortiori* of

$$\sum_r c_r k^r q_{r,k}$$

and hence also of

$$\sum_{\nu=1}^{\nu=N} \sum_r c_r k^r_\nu q_{r,k} e^{|k_\nu|}.$$

Since the coefficients C are finite in value and in number, the sum which occurs in the second line of (2) can, by increasing p, be made as small as we please.

The numbers of the first line are, after the first, all divisible by p (3), but the first number, $C_0\psi(h)$, is not (1). Therefore the sum of the numbers in the first line is not divisible by p and hence is different from zero. The sum of an integer and a fraction cannot be zero. Hence equation (2) is impossible and consequently also equation (1).*

4. We now come to a proposition more general than the preceding, but whose demonstration is an immediate conse-

* The proof for the more general case where $C_0 = 0$ may be reduced to this by multiplication by a suitable factor, or may be obtained directly by a proper modification of $\psi(h)$.

quence of the latter. For this reason we shall call it Lindemann's corollary.

The number e *cannot satisfy an equation of the form*

$$(3)\quad C'_0 + C'_1 e^{k_1} + C'_2 e^{l_1} + \cdots = 0,$$

in which the coefficients are integers even when the exponents $k_1, l_1, \cdots$ *are unrelated algebraic numbers.*

To demonstrate this, let $k_2, k_3, \cdots, k_N$ be the other roots of the equation satisfied by k_1; similarly for $l_2, l_3, \cdots, l_{N'}$, etc. Form all the polynomials which may be deduced from (3) by replacing k_1 in succession by the associated roots $k_2, \cdots$, l_1 by the associated roots $l_2, \cdots$ Multiplying the expressions thus formed we have the product

$$\prod_{\alpha, \beta, \cdots} \{C'_0 + C'_1 e^{k_\alpha} + C'_2 e^{l_\beta} + \cdots\} \qquad \begin{bmatrix} \alpha = 1, 2, \cdots, N \\ \beta = 1, 2, \cdots, N' \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \end{bmatrix}$$

$$= C_0 + C_1(e^{k_1} + e^{k_2} + \cdots + e^{k_N}) + C_2(e^{k_1+k_2} + e^{k_2+k_3} + \cdots) + C_3(e^{k_1+l_1} + e^{k_1+l_2} + \cdots) + \cdots$$

In each parenthesis the exponents are formed symmetrically from the quantities $k_i, l_i, \cdots$, and are therefore roots of an algebraic equation with integral coefficients. Our product comes under Lindemann's theorem; hence it cannot be zero. Consequently none of its factors can be zero and the corollary is demonstrated.

We may now deduce a still more general theorem.

The number e *cannot satisfy an equation of the form*

$$C_0^{(1)} + C_1^{(1)} e^{k} + C_2^{(1)} e^{l} + \cdots = 0$$

where the coefficients as well as the exponents are unrelated algebraic numbers.

For, let us form all the polynomials which we can deduce from the preceding when for each of the expressions $C^{(1)}_i$ we substitute one of the associated algebraic numbers

$$C_i^{(2)},\ C_i^{(3)},\ \cdots\ C_i^{(N)}.$$

If we multiply the polynomials thus formed together we get the product

$$\prod_{\alpha, \beta, \gamma, \cdots} \{C_0^{(\alpha)} + C_1^{(\beta)}e^k + C_2^{(\gamma)}e^l + \cdots\} \quad \begin{bmatrix} \alpha = 1, 2, \cdots, N_0 \\ \beta = 1, 2, \cdots, N_1 \\ \gamma = 1, 2, \cdots, N_2 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \end{bmatrix}$$

$$\begin{aligned} = C_0 &+ C_k e^k + C_l e^l + \cdots \\ &+ C_{k,k} e^{k+k} + C_{k,l} e^{k+l} + \cdots \\ &+ \cdots \cdots \cdots \\ &+ \cdots \cdots \cdots, \end{aligned}$$

where the coefficients C are integral symmetric functions of the quantities

$$\begin{matrix} C_0^{(1)}, & C_0^{(2)}, & \cdots, & C_0^{(N_0)}, \\ C_1^{(1)}, & C_1^{(2)}, & \cdots, & C_1^{(N_1)}, \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{matrix}$$

and hence are rational. By the previous proof such an expression cannot vanish, and we have accordingly Lindemann's corollary in its most general form:

The number e *cannot satisfy an equation of the form*

$$C_0 + C_1 e^k + C_2 e^l + \cdots = 0$$

where the exponents k, l, $\cdots$ *as well as the coefficients* C_0, C_1, $\cdots$ *are algebraic numbers.*

This may also be stated as follows:

In an equation of the form

$$C_0 + C_1 e^k + C_2 e^l + \cdots = 0$$

the exponents and coefficients cannot all be algebraic numbers.

5. From Lindemann's corollary we may deduce a number of interesting results. First, the transcendence of π is an immediate consequence. For consider the remarkable equation

$$1 + e^{i\pi} = 0.$$

The coefficients of this equation are algebraic; hence the exponent $i\pi$ is not. Therefore, π is transcendental.

6. Again consider the function $y = e^x$. We know that $1 = e^0$. This seems to be contrary to our theorems about the transcendence of e. This is not the case, however. We must notice that the case of the exponent 0 was implicitly excluded. For the exponent 0 the function $\psi(x)$ would lose its essential properties and obviously our conclusions would not hold.

Excluding then the special case ($x = 0$, $y = 1$), Lindemann's corollary shows that in the equation $y = e^x$ or $x = \log_e y$, y and x, *i.e.*, the number and its natural logarithm, cannot be algebraic simultaneously. To an algebraic value of x corresponds a transcendental value of y, and conversely. This is certainly a very remarkable property.

If we construct the curve $y = e^x$ and mark all the algebraic points of the plane, *i.e.*, all points whose coördinates are algebraic numbers, the curve passes among them without meeting a single one except the point $x = 0$, $y = 1$. The theorem still holds even when x and y take arbitrary complex values. The exponential curve is then transcendental in a far higher sense than ordinarily supposed.

7. A further consequence of Lindemann's corollary is the transcendence, in the same higher sense, of the function $y = \sin^{-1} x$ and similar functions.

The function $y = \sin^{-1} x$ is defined by the equation

$$2ix = e^{iy} - e^{-iy}.$$

We see, therefore, that here also x and y cannot be algebraic simultaneously, excluding, of course, the values $x = 0$, $y = 0$. We may then enunciate the proposition in geometric form:

The curve $y = \sin^{-1} x$, *like the curve* $y = e^x$, *passes through no algebraic point of the plane, except* $x = 0$, $y = 0$.

CHAPTER V.

The Integraph and the Geometric Construction of π.

1. Lindemann's theorem demonstrates the transcendence of π, and thus is shown the impossibility of solving the old problem of the quadrature of the circle, not only in the sense understood by the ancients but in a far more general manner. It is not only impossible to construct π with straight edge and compasses, but there is not even a curve of higher order defined by an integral algebraic equation for which π is the ordinate corresponding to a rational value of the abscissa. An actual construction of π can then be effected only by the aid of a transcendental curve. If such a construction is desired, we must use besides straight edge and compasses a "transcendental" apparatus which shall trace the curve by continuous motion.

2. Such an apparatus is the *integraph*, recently invented and described by a Russian engineer, Abdank-Abakanowicz, and constructed by Coradi of Zürich.

This instrument enables us to trace the *integral curve*

$$Y = F(x) = \int f(x)\,dx$$

when we have given the *differential curve*

$$y = f(x).$$

For this purpose, we move the linkwork of the integraph so that the *guiding point* follows the differential curve; the *tracing point* will then trace the integral curve. For a fuller description of this ingenious instrument we refer to the original memoir (in German, Teubner, 1889; in French, Gauthier-Villars, 1889).

We shall simply indicate the principles of its working. For any point (x, y) of the differential curve construct the auxiliary triangle having for vertices the points (x, y), $(x, 0)$, $(x-1, 0)$; the hypotenuse of this right-angled triangle makes with the axis of X an angle whose tangent $= y$.

Hence, *this hypotenuse is parallel to the tangent to the integral curve at the point* (X, Y) *corresponding to the point* (x, y).

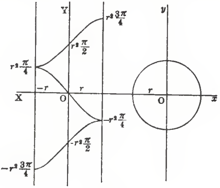


FIG. 16.

The apparatus should be so constructed then that the tracing point shall move parallel to the variable direction of this hypotenuse, while the guiding point describes the differential curve. This is effected by connecting the tracing point with a sharp-edged roller whose plane is vertical and moves so as to be always parallel to this hypotenuse. A weight presses this roller firmly upon the paper so that its point of contact can advance only in the plane of the roller.

The practical object of the integraph is the approximate evaluation of definite integrals; for us its application to the construction of π is of especial interest.

3. Take for differential curve the circle

$$x^2 + y^2 = r^2;$$

the integral curve is then

$$Y = \int \sqrt{r^2 - x^2}dx = \frac{r^2}{2}\sin^{-1}\frac{x}{r} + \frac{x}{2}\sqrt{r^2 - x^2}.$$

This curve consists of a series of congruent branches. The points where it meets the axis of Y have for ordinates

$$0, \quad \pm\frac{r^2\pi}{2}, \quad \cdots$$

Upon the lines $X = \pm r$ the intersections have for ordinates

$$r^2\frac{\pi}{4}, \quad r^2\frac{3\pi}{4}, \quad \cdots$$

If we make $r = 1$, the ordinates of these intersections will determine the number π or its multiples.

It is worthy of notice that our apparatus enables us to trace the curve not in a tedious and inaccurate manner, but with ease and sharpness, especially if we use a tracing pen instead of a pencil.

Thus we have an actual constructive quadrature of the circle along the lines laid down by the ancients, for our curve is only a modification of the quadratrix considere by them.

NOTES

PART I — CHAPTER III

Gaussian Polygons. Up to the time of Gauss, no one suspected that it was possible to construct, with ruler and compasses, regular polygons other than those the number of whose sides could be expressed in one of the forms: 2^n, $2^n \cdot 3$, $2^n \cdot 5$, $2^n \cdot 15$. All of these were known to the Greeks. But Gauss proved as early as 1801[1] that whenever a prime number F_μ could be expressed in the form $2^{2^\mu}+1$, the construction of a regular polygon with F_μ sides was possible by Euclidean methods. It was then apparent that regular polygons not included in the Euclidean series, namely 17, 257, 65537, . . . sides, could be constructed under the same imposed conditions. And indeed Gauss's discussion led to the result[2], that the *only* regular polygons which it is possible to construct with ruler and compasses, are those the number P of whose sides can be expressed in the form

$$2^\alpha \cdot (2^{2^{\alpha_1}} + 1) \cdot (2^{2^{\alpha_2}} + 1) \cdot (2^{2^{\alpha_3}} + 1) \cdots (2^{2^{\alpha_s}} + 1),$$

where $\alpha\ \ldots\ \alpha_s$ are distinct positive integers and each $2^{2^{\alpha_i}}+1$ is a prime. The number of such polygons is small in comparison with the number of regular polygons which can not be constructed with the means employed. As Dickson has pointed out[3] the number of P's up to 100 is 24; up to 300 is 37 (all noted by Gauss); up to 1000 is 52; up to 1000000 only 206. Kraitchik has remarked[4] that there are only 30 polygons with an odd number of sides that are known to be constructible with ruler and compasses. These polygons have the following number of sides: 5, 15, 17, 51, 85, 255, 257, 771, 1285, 3855, 4369,13107, 21845, 65535, 65537, 196611, 327685, 983055, 1114129, 3342387, 5570645, 16711935, 16843009, 50529027, 84215045, 252645135, 286331153, 858993459, 1431655765,

---

[1] *Disquistiones arithmeticæ*, Leipzig, 1801, p. 664; *Werke*, v. 1., 2. Abdruck, 1870, p. 462: French ed. *Recherches Arithmétiques*, Paris, 1807, p. 488; Ger. ed. by Maser, Berlin, 1889, p. 447.

[2] This result was, in effect, stated, but not proved, by Gauss.

[3] L. E. Dickson, "On the number of inscriptible regular polygons", *Bull. N. Y. Math. Soc.*, Feb., 1894, v. 3, p. 123.

[4] Kraitchik, *Recherches sur la théorie des nombres*, Paris, 1924, p. 270.

4294967295. This set of numbers, together with 1 and 3, coincides. with the divisors of $2^{32} - 1 = 1 \cdot 3 \cdot 5 \cdot 17 \cdot 257 \cdot 65537$.

The determination of the number of regular polygons which can be constructed for P less than a given integer is, then, bound up in the determination of the prime numbers F_μ. Now for only 18 values of μ has it been shown whether F_μ is prime or not, namely for the values of μ from 0 to 9 inclusive, and for 11, 12, 15, 18, 23, 36, 38, 73. In the first five of these cases, and in these alone, is F_μ prime. These five cases were noted by Fermat in the seventeenth century. It may well turn out that F_μ is not prime, for $\mu > 4$, although Eisenstein proposed as a problem[1]: "There are an infinity of prime numbers of the form $2^{2^\mu} + 1$".

The results already established in this connection may be set forth in tabular form[2]:

μ	Prime Factors of $F_\mu = 2^{2^\mu} + 1$	Discoverer	Year of Discovery
0-4	3, 5, 17, 257, 65537	Fermat	1640
5	$\begin{cases} 2^7 \cdot 5 + 1 = 641 \\ 2^7 \cdot 52347 + 1 = 6700417 \end{cases}$	L. Euler	1732
6	Unknown bnt composite.	Lucas	1878
	$2^8 \cdot 9 \cdot 7 \cdot 17 + 1 = 274177$	Landry	1880
	$2^8 \cdot 5 \cdot 5256 2829149 + 1 + 67280421310721$	Landry and Le Lasseur	1880
7	Unknown hut composite.	A. E Western, J.C, Morehead	1905
8	Unknown hut composite.	A. E. Western J. C. Morehead	1909
9	$2^{16} \cdot 37 + 1 = 2424833$	A. E. Western	1903
11	$\begin{cases} 2^{13} \cdot 3 \cdot 13 + 1 = 319489 \\ 2^{13} \cdot 7 \cdot 17 + 1 = 974849 \end{cases}$	A, Cunningham	1899
12	$2^{14} \cdot 7 + 1 = 114689$	E. A. Lncas and P. Pervouchine	1877
	$\begin{cases} 2^{16} \cdot 397 + 1 = 26017793 \\ 2^{16} \cdot 7 \cdot 139 + 1 = 63766529 \end{cases}$	A. E. Western	1903
15	$2^{21} \cdot 579 + 1 = 1214251009$	M. Kraitchik	1925
18	$2^{20} \cdot 13 + 1 = 13631489$	A. E. Western	1903
23	$2^{25} \cdot 5 + 1 = 167772161$	P. Pervouchine	1878
36	$2^{39} \cdot 5 + 1 = 2748779069441$	Seelhoff	1886
38	$2^{41} \cdot 3 + 1 = 6597069766657$	J. Cullen, A. Cunningham, A. E. and F. J. Western	1903
73	$2^{75} \cdot 5 + 1 = 188894559314785808547841$	J. C. Morehead	1906

[1] G. Eisenstein, "Aufgaben", Crelle's *Journal*, v. 27, 1844, p. 87.

[2] The sources for the different results, except those of Fermat, are as follows, for the 13 different values of μ:

The labor expended in deriving these results has been enormous; to the layman who knows nothing of congruences in the theory of numbers, the facts found must seem almost to border on the miraculous. For, even when $\mu = 10$, a case not yet solved, F_μ contains 309 digits; but when $\mu = 36$, F_μ is a number of more than twenty trillion digits. Concerning it Lucas remarked[1] "la bande de papier qui le contiendrait ferait le tour de la Terre". For $\mu = 73$, Ball states that the digits in F_μ "are so numerous that, if the number were printed in full with the type and number of pages used in this book [*Mathematical Recreations*, fifth edition, 1911, 508 pages], many more

---

5. L. Euler, *Commentarii Academiæ Scientiarium Petrop.*, v. 6 (1732—3), 1738, p. 104; laid hefore the Academy of St. Petershurg, 26. Sept. 1732.
In his autohiography (Springfield, Mass., 1833, p. 38) the American calculator Zera Colhurn records that while on exhibition in London, at the age of 8, he found "hy the mere operation of his mind" the factors 641 and 6, 700, 417 of 4, 294, 967, 297 ($= 2^{32} + 1$). *Cf.* F. D. Mitchell, "Mathematical prodigies", *Amer. Journal of Psychology*, v. 18, 1907, p. 65.

6. Lucas, *Comptes Rendus de l'Académie des Sciences*, Paris, v. 85, 1878, p. 138; *Amer. Jour. Math.*, v. 1, 1878, p. 238; *Recreations mathématiques*, v. 2 (2e éd., 1896), p. 234—5. Landry, *Nouv. Corresp. Math.*, v. 6, 1880, p. 417.

7. Independent discoverers: Western, *Proc. Lond. Math. Soc.*, s. 2, v. 3, p. xxi—xxii. Ahstract of paper read, April 13, 1905; Morehead, *Bull. Amer. Math. Soc.*, v. 11, p. 543—545, ahstract of paper read April 29, 1905.

8. Western and Morehead, *Bull. Amer. Math. Soc.*, v. 16, 1909, p. 1—6; "each doing half of the whole work".

9, 12 (Western), 13, 16. *Proc. Lond. Math. Soc.*, s. 2, v. 1, 1903, p. 175; abstract of paper read May 14, 1903.

11. A. Cunningham, *Brit. Assoc. Rept.*, 1899, p. 653—4; the factors are here given as 319489 and 974489. The second number is incorrect, 4 and 8 heing interchanged. The other forms of the correct factors were given by A. Cunningham and A. E. Western in *Proc. Lond. Math. Soc.*, s. 2, v. 1, 1903, p. 175. It is here noted also that there are no more factors of $F\mu < 10^6$, and no other factor of $F\mu < 10^6$, (μ not less than 14).

12, 23. E. Lucas, *Atti Accad. Torino*, v. 13 (1877—8), p. 271 [27 Jan., 1878]. *Mélanges math. ast. acad. Pétersb.*, v. 5, part 5, 1879, p. 505, 519, or *Bull. Acad. Pétersb.*, s. 3, v. 24, 1878, p. 559; s. 3, v. 25, 1879, p. 63; communication of results, for $\mu = 12$ and 23, found hy J. Pervouchine, in Nov. 1877 and Jan. 1878. He notes that the integer $2^{2^{23}} + 1$ contains 2525223 digits.

15. M. Kraitchik, *Comptes Rendus de l'Académie des Sciences*, Paris, v. 180, p. 800, March, 1925; also *Sphinx-Oedipe*, v. 20, p. 24.

36. P. Seelhoff, *Zeitschrift math. u. Phys.*, v. 31, 1886, p. 174.

73. J. C. Morehead, *Bull. Amer. Math. Soc.*, v. 12, 1906, p. 449—451.

[1] E. Lucas, *Théorie des nombres*, Paris, v. 1, 1891, p. 51.

volumes would be required than are contained in all the public libraries of the world".

In not less than seven places[1], during the years 1640–58, did Fermat refer to $F_\mu = 2^{2^\mu} + 1$ as representing a series of prime numbers; but in no place did he claim that F_μ was always prime.

Gauss's Statement of his Polygon Results. In two passages the implication to be drawn from what Klein has written is, that Gauss published a proof that a regular polygon of p sides can not be constructed by ruler and compasses if p is a prime not of the form $2^k + 1$. The passages to which I refer are (pages 2, 16):

(1) "Gauss added other cases [to Euclid's] by showing the possibility of the division into parts where p is a prime number of the form $p = 2^{2^\mu} + 1$, and the impossibility for all other numbers"; (2) "Gauss extended this series of numbers [Euclid's] by showing that the division is possible for every prime number of the form $p = 2^{2^\mu} + 1$ but impossible for all other prime numbers and their powers". Now the implication referred to above is not correct, as Pierpont interestingly set forth in his paper "On an undemonstrated theorem of the *Disquisitiones Arithmeticae*"[2]. That is, Gauss *did not give a proof* of the "impossibility" referred to in the quotations. But after proving the "possibility" described above he continued as follows:

"As often as $p-1$ contains other prime factors besides 2, we errive at higher equations[3], namely, to one or more cubic equationas, if 3 enters

[1] Letter dated Aug. [?] 1640 to Frenicle (*Oeuvres de Fermat*, v. 2, 1894, p. 206); letter dated 18 Oct., 1640, to Frenicle (*Oeuvres*, v. 2, 1894, p. 208); *Varia Opera*, Toulouse, 1679, p. 162; Brassine's *Précis*, Toulouse, 1853, p. 142—3); letter dated 25 Dec., 1640, to Mersenne (*Oeuvres*, v. 2, p. 212—213); "De solutione problematum geometriconum per curvas simplicissimas et unicuique problematum generi proprie convenientes, Dissertatio tripartita" (*Oeuvres de Fermat*, v. 1, 1891, p. 130—131; French translation, v. 3, 1896, p. 120; *Varia Opera*, 1679 [reprint, 1861], p. 115); letter dated 29 August, 1654, to Pascal (*Oeuvres de Pascal*, v. 4, Paris, 1819, p. 384; *Oeuvres de Fermat*, v. 2, 1894, p. 309—310); letter to Sir Kenelm Digby, sent by Digby to Wallis, 19 June, 1658 (*Oeuvres de Fermat*, v. 2, 1894, p. 402, 404—5; French translation of the Latin, v. 3, 1896, p. 314, 316); letter dated August, 1659 to Carcavi, copy sent by Carcavi to Huygens 14 August, 1659 (*Corresp.* de Huygens no. 651; *Oeuvres de Fermat*, v. 2, p. 433—434).

[2] *Bull. Amer. Math. Soc.*, v. 2, 1895, p. 77—83.

[3] In his earlier discussion of an inscribed polygon of p sides, Gauss considers the equation $x^p - 1 = 0$ and the resulting equation got by dividing out the factor $x - 1$, where p is a prime.

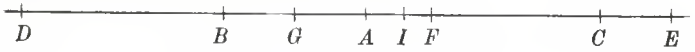
once or oftener as a factor of $p-1$, to equations of 5th degree if $p-1$ is divisible by 5, etc. And we can prove with all rigour that these equations cannot be avoided or made to depend upon equations of lower degree; and although the limits of this work do not permit us to give the demonstration here, we still thought it necessary to signal this fact in order that one should not seek to construct other polygons than those given by our theory, as, for example, polygons of 7, 11, 13, 19 sides, and so employ one's time in vain."

Fermat's Theorem. This theorem (p. 17) was indicated by Fermat in a letter, dated 18 October 1640, to B. Frenicle de Bessy (*Oeuvres de Fermat*, v. 2, 1894, p. 209). Euler gave two proofs (*Comment. Acad. Petrop.*, v. 8 for 1736, 1741, p. 141, and *Comment. Nov. Acad. Petrop.*, v. 7 for 1758-59, 1761, p. 49). Other proofs are due to Lagrange (*Nouv. Mém. de l'Acad. de Berlin*, 1771) and to Gauss (*Disquisitiones Arithmeticæ*, § 49)

PART I — CHAPTER IV

Geometrical Constructions of the Regular Heptadecagon. The remark of Klein (p. 24, 32) that we posses as yet no method of construction of the regular polygon of seventeen sides, based upon considerations purely geometrical, is a little curious, since several constructions of this kind have been given. One by Erchinger was indeed reported by Gauss in 1825[1]. The construction is as follows:

Let D, B, G, A, I, F, C, E be points on a line determined by constructions about to be given. Let AB be a line of any length. Produce it both ways to C and D so that,



$$AC \times BC = AB \times BD = 4\,\overline{AB}^2.$$

[1] *Gottingische gelehrte Anzeigen*, Dec. 19, 1825, no. 203, p. 2025; *Werke*, v. 2, p. 186–7. To Art. 365 of the *Disquisitiones Arithmeticae* Gauss added this note in his handwriting: "Circulum in 17 partes divisibilem esse geometrice, deteximus 1796 Mart. 30". *Cf. Werke* v. 1, p. 476 and v. 10_1, 1917, p. 3–4, 120–126, 488. The discovery of the result was first announced in the *Intelligenzblatt* of the *Allgemeine Literatur-Zeitung*, no. 66, 1 June, 1796, col. 554.

Further determine the points, E, G, on both sides of CA produced so that,

$$AE \times EC = AG \times CG = \overline{AB}^2;$$

and find the point F on the side A of the line BA produced, such that

$$AF \times DF = \overline{AB}^2.$$

Finally divide AE in I so that

$$AI \times EI = AB \times AF,$$

where AI is the smaller, and EI the larger part of AE. Then construct a triangle, in which each of two sides equals AB, the third being equal to AI. About this triangle describe a circle; then AI will be one side of the regular inscribed polygon of seventeen sides.

Gauss particularly remarks that the author gave a purely synthetic proof of this construction.

Another synthetic construction and proof dated "Dublin, 17th October, 1819" was published by Samuel James in the *Transactions of the Irish Academy*[1]. Yet another construction was given by John Lowry in *The Mathematical Repository*[2] for 1819. But the earliest published geometrical construction was given by Huguenin in his *Mathematische Beiträge zur weiteren Ausbildung angehender Geometer*, Königsberg, 1803, p. 283.

A score of geometrical constructions are assembled in A. Goldenring, *Die elementargeometrischen Konstruktionen des regelmässigen Siebzehnecks*, Leipzig, 1915. See also the review of this work in *Bull. Amer. Math. Soc.*, v. 22, 1916, p. 239—246, and my note "Gauss and the regular polygon of seventeen sides" in *Amer. Math. Monthly*, v. 27, 1920, p. 323—326.

The discovery that the regular polygon of seventeen sides could be constructed with ruler and compasses was not only one of which Gauss was vastly proud throughout his life, but also, according to Sartorius von Waltershausen[3], the one which decided him to dedicate his life to the study of mathematics. Archimedes expressed the wish that a sphere inscribed in a cylinder be inscribed on his tomb, as Ludolf van Ceulen did in connection with the value of π to 35 decimal

[1] V. 13 (1818), p. 175—187; paper read Jan. 24, 1820.

[2] N. s., v. 4, p. 160. Lowry's proof occupies p. 160—168.

[3] *Gauss zum Gedächtniss*, Leipzig, 1856, p. 16.

places, and Jacques Bernoulli with reference to the logarithmic spiral. So also, according to Weber[1], Gauss requested that the regular polygon of seventeen sides should be engraved on his tombstone. While this request was not granted, as it was in each of the other cases mentioned, it is engraved on the side of a monument to Gauss in Braunschweig, his birthplace.

Constructions in general with Ruler and Compasses. Regarding constructions as effected when intersections of circles with circles or lines, or of lines with lines may be determined, it can be shown that: *Every problem solved with ruler and compasses can be solved with compasses alone.* This was first shown by Georg Mohr in his *Euclides Danicus* published at Amsterdam in 1672; this work was reprinted in 1928 by the Danish Society of Sciences. Klein refers (p. 33) only to Mascheroni's proof of this result 125 years later, in his *Geometria del Compasso.* Of this work there were two French editions *Géométrie du Compas*, Paris, 1798 and 1828. From the first of these a German edition *L. Mascheroni's Gebrauch des Zirkels*, Berlin, 1825, was prepared by J. P. Gruson. The subject is treated in English by: A. Cayley, *Messenger of Math.*, v. 14, 1885, p.179—181; *Collected Papers*, v. 12, p. 314—317; by E. W. Hobson, in a presidential address, *Mathematical Gazette*, v. 7, 1913, p. 49—54; by H. P. Hudson, *Ruler & Compasses*, London, 1916, p. 131—143; and by J. Coolidge, *Treatise on the Circle and Sphere*, Oxford, 1916, p. 186—188.

Klein has noted (p. 33—34) that Poncelet first conceived the result that *given a circle and its center, every solution of a problem with ruler and compasses can be carried through with ruler alone.* A little later Klein states (p. 34) "we will show *how with the straight edge and one fixed circle we can solve every quadratic equation*". This is not possible; Klein should have had "with its center" after "one fixed circle". That the center be also given is very essential when only one circle is given. Hilbert suggested the problem: How many given circles in a plane are necessary in order to determine with ruler alone, the center of one of them? In 1912 D. Cauer[2] showed: (a) If two circles do not intersect in real points it is generally impossible to determine the center of either circle with ruler alone; (b) A center

---

[1] *Encyclopädie der elementaren Algebra und Analysis* bearbeitet von H. Weber. 2. ed. Leipzig, 1906, p. 362.

[2] *Mathematische Annalen*, v. 73, 1912, p. 90—94; v. 74, 1913, p. 462—464.

may be determined if the circles cut in real points, touch, or are concentric. About the same time J. Grossmann discovered a result which proved that *Every problem solvable with ruler and compasses can also be solved with ruler alone if we are given, in the plane of construction, three linearly independent circles.* Correct proofs of this result were given by Schur and Mierendorff.

From this it is clear that every construction with ruler and compasses can be effected with a ruler, and compasses with a fixed opening. Constructions of this kind were found already in the tenth century by Abû'l Wefâ of Bagdad[1]. With such means, in the sixteenth century, certain problems of Euclid were solved by Cardano, Ferraro, and Tartaglia. At Venice in 1553 G. B. Benedetti published a little treatise, *Resolutio omnium Euclidis problematum, aliorumque ad hoc necessario inventorum, una tantumodo circuli data apertura.* In English the topic is treated in a rare pamphlet translated from the Dutch by Joseph Moxon[2], and in an article by J. S. Mackay[3].

Every problem whose solution is possible by ruler and compasses can be also solved with a two edged ruler alone, whether the edges are parallel or meet in a point. For some of the literature in this connection the following sources may be consulted: *Nouvelle Corresp. Math.*, v. 3, 1877, p. 204—208; v. 5, 1879, p. 439—442; v. 6, 1880, p. 34—35; Akademie der Wissen., Vienna, *Sitzungsberichte*, Abt. IIa, v. 99, 1890, p. 854—858; *Bolletino di Matematiche e di Scienze fisiche e naturali*, v. 2, 1900—01, p. 129—145, 225—237.

PART II — CHAPTER II

Irrationality of π. Klein wrote (p. 59): "After 1770 critical rigour gradually began to resume its rightful place. In this year appeared the work of Lambert: *Vorläufige Kenntnisse für die, so die Quadratur*

[1] "Woepcke" Analyse et extrait d'un recueil de constructions géométriques par Aboûl Wafâ", *Journal Asiatique*, 1855.

[2] *Compendium Euclidis Curiosi: or, geometrical operations. Showing how with a single opening of the Compasses and a straight ruler all the propositions of Euclid's first five books are performed.* London, 1677. Moxon does not tell us who the author of the Dutch treatise was.

[3] "Solutions of Euclid's problems, with a ruler and one fixed aperture of the compasses, by the Italian geometers of the sixteenth century" *Proc. Edinb. Math. Soc.*, v. 5, 1887, p. 2—22.

. . . *des Cirkuls suchen.* Among other matters the irrationality of π is discussed. In 1794 Legendre in his *Éléments de Géométrie* showed conclusively that π and π^2 are irrational numbers." The implication of this note is that Lambert did not discuss the irrationality of π conclusively and that Legendre did. How both of these points of view are essentially incorrect will appear in what follows. Klein was simply reproducing the erroneous statements of Rudio[1]; but after Pringsheim's careful study in 1898[2], Lambert's proof emerged as "ausserordentlich scharfsinnig und im wesentlichen vollkommen einwandfrei", while Legendre's remained "in Bezug auf Strenge hinter Lambert weit zurück".

As in the later proof of the transcendence of π, so here when its irrationality was in question, discussion of e is fundamental. The irrationality of e and e^2 was shown, substantially, by Euler in 1737[3] and he gave the expression for e as a continued fraction on which Lambert's proofs of the irrationality of e^x, tan x and π rest. Starting with Euler's development[4]

$$\frac{e-1}{2} = \frac{1}{1+}\,\frac{1}{6+}\,\frac{1}{10+}\,\frac{1}{14+}\,\frac{1}{18+}\text{etc.},$$

Lambert found

$$\frac{e^x-1}{e^x+1} = \frac{1}{2/x+}\,\frac{1}{6/x+}\,\frac{1}{10/x+}\,\frac{1}{14/x+}\ etc.,$$

and since

[1] F. Rudio: *Archimedes, Huygens, Lambert, Legendre, vier Abhandlungen über die Kreismessung.* Leipzig, 1892, p. 56f. This error is also reproduced by B. Calò in Enriques's *Fragen der Elementargeometrie,* II. Teil, 1907, p. 315; by D. E. Smith in Young's *Monographs on Topics of Modern Mathematics,* 1911, p. 401. The matter was correctly set forth by T. Vahlen in *Konstruktionen und Approximationen,* Leipzig, 1911, p. 319f.

[2] A. Pringsheim: "Über die ersten Beweise der Irrationalität von e und π", Bayerische Akad. der Wissen., *Sitzungsberichte,* mathem.-phys. Cl., v. 28, 1899, p. 325–337.

[3] "De fractionibus continuis", *Comment. acad. de Petrop,* v. 9, 1744, p. 108. Presented to St. Petersburg Academy, March. 1737.

[4] L. Euler: *Introductio in analysin infinitorum.* Tomus Primus, Lausannae, 1748, p. 319. This work was finished in 1745; *Cf.* G. Eneström, *Verzeichnis* etc., Erste Lieferung, p. 25.

$$\frac{e^x - 1}{e^x + 1} = \frac{e^{x/2} - e^{-x/2}}{e^{x/2} + e^{-x/2}} = \tanh \frac{x}{2} = \frac{1}{i} \tan \frac{ix}{2}, \text{ if } z = \frac{ix}{2},$$

$$\tan z = \frac{1}{1/z -} \frac{1}{3/z -} \frac{1}{5/z -} \frac{1}{7/z -} \frac{1}{9/z -} \cdots .$$

He then proved the theorems:

1. *If x is a rational number different from zero, e^x can never be rational.*

For $x = 1$, we have as special case the irrationality of e.

2. *If z is a rational number different from zero, tan z can never be rational.*

For $z = \pi/4$, $\tan \pi/4 = 1$, and hence as a special case the irrationality of π.

The part of Lambert's *Vorläufige Kenntnisse* to which Klein refers contains some formulae without proof, and no analytical developments, and was rather intended to serve as a popular survey of the treatment of the topic. With it must be considered the scientifically remarkable "*Mémoire*" of 1767[1]. Here "mit minutiöser Genauigkeit" Lambert proves the convergence of the expression for tan z as a continued fraction. Pringsheim dwells on the "astounding" nature of these considerations at this period in the history of mathematical thought. For of such considerations Legendre was innocent, as well as the great Gauss in his 1812 memoir on hypergeometric series, and others, till a much later period.

"Thus the Lambert memoir contains the *first*, and for many years, the *only* example of what we now consider really rigorous developments of functions as converging continued fractions, in particular, that for tan z given above."

Measurement of a Circle. By considering inscribed and circumscribed polygons up to 96 sides Archimedes arrived at the result that the ratio of the circumference of a circle to its diameter is less than $3\frac{10}{70}$ but greater than $3\frac{10}{71}$. The following table exhibits the perimeters of regular inscribed and circumscribed polygons of a circle with a unit diameter (Chauvenet, *Treatise on Elementary Geometry*, Philadelphia, 1870, p. 161).

[1] "Mémoire sur quelques propriétés remarquables des quantités transcendantes circulaires et logarithmiques". Lu en 1767. Printed in 1768 in *Hist. de l'acad. royale des sciences et belles-lettres*, Berlin, Année 1761 (!), p. 265—322.

Number of sides	Perimeter of circumscribed polygon	Perimeter of inscribed polygon
4	4.0000000	2.8284271
8	3.3137085	3.0614675
16	3.1825979	3.1214452
32	3.1517249	3.1365485
64	3.1441184	3.1403312
128	3.1422236	3.1412773
256	3.1417504	3.1415138
512	3.1416321	3.1415729
1024	3.1416025	3.1415877
2048	3.1415951	3.1415914
4096	3.1415933	3,1415923
8192	3.1415928	3.1415926

The remarkable approximation 355/113 for π is correct to six places of decimals. It seems to have been first given by a Chinese, Tsu Ch'ung-ching (5th century), and later by Valentin Otho (16th century) and Adriaen Anthonisz (17th century). Grunert gave a geometrical construction for π based on the fact that $355/113 = 3 + 4^2/(7^2 + 8^2)$, *Archiv der Mathematik und Physik*, v. 12, 1849, p. 98.
Another construction was given by Ramanujan in *Journ. Indian Math. Soc.*, v. 5, 1913, p. 132 (also in *Collected Papers of Srinivasa Ramanujan*, Cambridge, 1927, p. 22, 35).

Euler's Formula. The formula

$$e^{ix} = \cos x + i \sin x$$

was first given by Euler in *Miscellanea Berolinensia*, v. 7, 1743, p. 179 (paper read 6 Sept. 1742), and again in his *Introductio in Analysin*, Lausanne, 1748, v. 1, p. 104. He gave also

$$e^{-ix} = \cos\ x - i \sin\ x.$$

The equivalent of the form

$$ix = \log\ (\cos\ x + i \sin\ x$$

was given earlier by Roger Cotes (*Philosophical Transactions*, 1714, v. 29, 1717, p. 32) as: "Si quadrantis circuli quilibet arcus, radio

CE descriptus, sinum habeat *CX*, sinumque complementi ad quadrantem *XE*: sumendo radium *CE*, pro Modulo, arcus erit rationis inter $EX + XC\sqrt{-1}$ & *CE* mensura ducta in $\sqrt{-1}$." See also Cotes, *Harmonia Mensurarum*, 1722, p. 28.

PART II — CHAPTER IV

In the course of the discussion on pages 61—74 it is assumed that there are an infinite number of prime numbers. One of the neatest proofs of this fact was given by Euclid (about 300 B.C.) in proposition 20, book 9 of his Elements.

On page 77, in considering the relation $y = e^x$, Klein made a slight slip when he wrote: "To an algebraic value of x corresponds a transcendental value of y, and conversely." "Conversely" leads to the statement, to a transcendental value of y corresponds an algebraic value of x. But a proof of this has nowhere been given; indeed the result is not true, in general. To correct delete "conversely" and add: "To an algebraic value of y corresponds a transcendental value of x."

FROM DETERMINANT TO TENSOR

BY

W. F. SHEPPARD, Sc.D., LL.M.
FORMERLY FELLOW OF TRINITY COLLEGE, CAMBRIDGE

PREFACE

THE tensor calculus used in the mathematical treatment of relativity, and concisely explained by Professor A. S. Eddington in his 'Report on the Relativity Theory of Gravitation', is, like the various kinds of vector calculus, a system of condensed notation which not only conduces to economy in the writing of symbols, but, what is more important, enables spatial and physical relationships to be grasped as a whole without having to be built up from a number of components which really represent views from different parts of space. Three-dimensional geometry or physics is troublesome enough: the addition of a fourth dimension made the need of a condensed notation imperative.

Professor Eddington has recently pointed out that the tensor notation and methods can be applied, with happy results, to other and more elementary classes of problems than those for which they were originally devised; and this book is an attempt to put his somewhat compressed exposition into a form in which it may appeal to a larger circle of readers. The book, therefore, is not intended as an introduction to the mathematical theory of relativity—though I hope it may be of some use for that purpose—but rather as an exercise in the elementary application of methods which, apart from any practical use, possess a special beauty of their own.

The new notation is not introduced until the fifth chapter. The properties of determinants, which serve as the starting-point for the application of the notation, are familiar to the mathematician; but, as I hope the book may be read by some who are not entirely at ease with determinants,

I have commenced with four chapters on the elementary theory of the subject. I make no apology for doing this, instead of referring the reader to the ordinary text-book on algebra. The text-book treatment is not always stimulating; the reasons for the various stages are not necessarily clear to the student; and attempt at simplicity sometimes leads to loss of rigidity in proof. In such a subject it is necessary to take the reader into one's confidence; and this earlier part may in this respect be found helpful to some, teachers or students, to whom the later part makes at first a less strong appeal.

I have added a chapter on some applications to the theory of statistics, to which the tensor calculus seems specially suitable. The basis of this portion, so far as method is concerned, is a short paper by Professor Eddington, mentioned at the end of the chapter. This is one only of many possible applications.

What I have called double sets will be recognized by the advanced student as matrices; and many of the propositions will be found to be familiar. But the tensor calculus may fairly claim that, in bringing into close relation various branches of mathematical study, previously regarded as distinct, it gives them a new life.

I have to thank Professor Eddington for looking at my manuscript and making some corrections and suggestions.

5 *June* 1923. W. F. S.

CONTENTS

INTRODUCTION

As this is a comparatively new subject to most readers, it may be as well to explain briefly what it is about.

A vector in (say) 3 dimensions is a directed quantity, determined as regards both direction and magnitude by its components, which are magnitudes measured along three definite axes. These axes being supposed to have been fixed beforehand, we can take them in some definite order; and a vector $\mathfrak{A}$ is then determined by a set of 3 quantities, which we may call

$$A_1 \quad A_2 \quad A_3.$$

Algebraically, the idea of a vector can be extended to something which is determined by a set of m quantities

$$A_1 \quad A_2 \quad A_3 \ldots A_m,$$

where m may have any value.

A determinant (say)

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

is the algebraical sum of all the products that can be formed in a certain way according to a certain rule of signs from the set of quantities

$$\begin{matrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3. \end{matrix}$$

Each 'element' of this set has its position fixed in the set by the numbers of horizontal and of vertical steps required to reach it from the initial element a_1. Thus the set is extended in two directions, while the set which determines a vector is extended in one direction only. This applies to a determinant with any number m^2 of elements.

In the same way we might have a set extended in 3 directions, the symbols being written along the edges of a cube and along lines inside the cube or on its faces; and we could, in theory, increase the number of 'directions' to 4 or more, by proper convention as to the order in which the elements are to be taken. On the other hand, a single quantity—what in the language of vectors is called a scalar—may be regarded as a set not extended in any direction.

The tensor calculus, using the word 'tensor' in its broad sense, deals with all these different kinds of sets, in relation to sets of variables by which we can regard axes of reference as being determined. In the narrower sense in which the word is used in reference to the theory of relativity, only sets which satisfy certain conditions are called tensors.

In this book I have treated the tensor calculus as arising out of the use of determinants. Chapters I–IV deal with the elementary theory of determinants, so far as it is required for our purpose. (The student who is familiar with determinants can skip these chapters.) In Chapter V the tensor notation is introduced in successive steps, with explanatory remarks. These latter are in small print, not because they are less important, but in order not to break the continuity of the chapter as a whole. In Chapter VI these explanatory paragraphs (or parts of them) are brought together and amplified so as to give a general idea of the elementary properties of sets. Chapters VII and VIII deal with some developments of the subject in its general aspect. Chapter IX shows the application of the methods to certain problems in the theory of statistics and of error; this can be omitted by any one who wishes to pass on to Chapter X, which deals very briefly with the tensor in its more limited sense, as applied to the theory of relativity.

DETERMINANTS

I. ORIGIN OF DETERMINANTS

I. 1. **Solution of simultaneous equations.**—Determinants ordinarily arise out of the solution of simultaneous equations. Suppose we have two equations

$$\left.\begin{aligned} 5x+2y &= 19 \\ 4x+3y &= 18 \end{aligned}\right\}.$$

Then, if we used only elementary methods, we could multiply the first by 3 and the second by 2, which would give

$$\left.\begin{aligned} 15x+6y &= 57 \\ 8x+6y &= 36 \end{aligned}\right\};$$

and thence, by subtracting, we should have

$$7x = 21, \quad x = 3,$$

whence either equation would give

$$y = 2.$$

Similarly, if we had three equations

$$\left.\begin{aligned} 2x+5y+3z &= 4 \\ x-3y-2z &= -1 \\ -5x-4y+z &= 7 \end{aligned}\right\},$$

we could, by eliminating z between the first and the second and between the second and the third, obtain

$$\left.\begin{aligned} 7x+y &= 5 \\ -9x-11y &= 13 \end{aligned}\right\};$$

whence, proceeding as before, we should obtain

$$x = 1, \quad y = -2, \quad z = 4.$$

This process of successive elimination is tedious, especially when there are more than two unknowns; and it is found

better to obtain a formula for the general solution and apply it to the numerical values of the particular case.

Since such an equation as $x-3y-2z = -1$ can be written in the form $(+1)x+(-3)y+(-2)z = (-1)$, we can use positive signs throughout, it being understood that the quantity represented by any symbol may be either positive or negative.

I. 2. **Formula for solution.**—(i) For completeness we begin with one unknown. The equation

$$a_1 x = k_1$$

gives

$$x = \frac{k_1}{a_1}.$$

(ii) For two unknowns the equations may be written

$$\left.\begin{aligned} a_1 x + b_1 y &= k_1 \\ a_2 x + b_2 y &= k_2 \end{aligned}\right\}.$$

Multiplying the first equation by b_2 and the second by b_1, and subtracting, we get

$$(a_1 b_2 - a_2 b_1) x = k_1 b_2 - k_2 b_1,$$

whence

$$x = \frac{k_1 b_2 - k_2 b_1}{a_1 b_2 - a_2 b_1}.$$

Similarly, interchanging a's and b's,

$$y = \frac{k_1 a_2 - k_2 a_1}{b_1 a_2 - b_2 a_1}$$

$$= \frac{a_1 k_2 - a_2 k_1}{a_1 b_2 - a_2 b_1}.$$

It should be noticed that the expressions for x and for y have the same denominator, and that the numerators are

obtained from the denominator by replacing the a's in the one case, and the b's in the other, by k's.

(iii) Next take the case of three unknowns. Let the equations be

$$\left.\begin{aligned} a_1x+b_1y+c_1z &= k_1 \\ a_2x+b_2y+c_2z &= k_2 \\ a_3x+b_3y+c_3z &= k_3 \end{aligned}\right\}.$$

Eliminating z from the first two equations, we obtain

$$(a_1c_2-a_2c_1)x+(b_1c_2-b_2c_1)y = k_1c_2-k_2c_1.$$

Similarly from the second and third equations

$$(a_2c_3-a_3c_2)x+(b_2c_3-b_3c_2)y = k_2c_3-k_3c_2.$$

Then, eliminating y from these equations, we get

$$x = \frac{(k_1c_2-k_2c_1)(b_2c_3-b_3c_2)-(k_2c_3-k_3c_2)(b_1c_2-b_2c_1)}{(a_1c_2-a_2c_1)(b_2c_3-b_3c_2)-(a_2c_3-a_3c_2)(b_1c_2-b_2c_1)}.$$

As before, the numerator is got from the denominator by replacing a's by k's, and we therefore need only consider the denominator. Multiplying out, it becomes

$$\begin{aligned} &a_1b_2c_2c_3-a_1b_3c_2^2-a_2b_2c_1c_3+a_2b_3c_1c_2-a_2b_1c_2c_3+a_2b_2c_1c_3 \\ &\qquad\qquad +a_3b_1c_2^2-a_3b_2c_1c_2 \\ &= c_2(a_1b_2c_3-a_1b_3c_2-a_2b_1c_3+a_2b_3c_1+a_3b_1c_2-a_3b_2c_1). \end{aligned}$$

Hence, replacing the a's by k's for the numerator,

$$x = \frac{k_1b_2c_3-k_1b_3c_2-k_2b_1c_3+k_2b_3c_1+k_3b_1c_2-k_3b_2c_1}{a_1b_2c_3-a_1b_3c_2-a_2b_1c_3+a_2b_3c_1+a_3b_1c_2-a_3b_2c_1}.$$

Corresponding expressions can be obtained for y and for z.

I. 3. **General problem.**—(i) We might proceed in the same way for equations in four or more unknowns. But this would mean that each case would have to be considered separately; and not only should we fail to get a general formula, but the algebraical work would soon become practically impossible. We therefore alter our tactics.

We write down the general equations involving m unknowns $x, y, z, \dots w$

$$\left.\begin{array}{l} a_1x+b_1y+\ c_1z+\dots+f_1w=k_1 \\ a_2x+b_2y+\ c_2z+\dots+f_2w=k_2 \\ \qquad\vdots \\ a_mx+b_my+c_mz+\dots+f_mw=k_m \end{array}\right\}, \quad \text{(I. 3. A)}$$

guess at a solution, and then verify that this solution does actually satisfy the equations.

(ii) The values of $x, y, z \dots w$ as found from these equations will be in the form of fractions. We will consider first the denominators. Putting together the results obtained in § 2, for the cases of $m = 1, 2, 3$, we find that the successive denominators, which we will call $D^{(1)}, D^{(2)}, D^{(3)}$, are

$$\left.\begin{array}{l} D^{(1)} = a_1 \\ D^{(2)} = a_1b_2-a_2b_1 \\ D^{(3)} = a_1b_2c_3-a_1b_3c_2-a_2b_1c_3+a_2b_3c_1+a_3b_1c_2 \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad -a_3b_2c_1 \end{array}\right\}. \quad \text{(I. 3. 1)}$$

We want to obtain an expression $D^{(m)}$, which we should guess to be the common denominator in the solution of the equations (I. 3. A), and of which $D^{(1)}, D^{(2)}, D^{(3)}$ shall be the particular cases for $m = 1, 2, 3$.

(iii) The three D's in (I. 3. 1) have a general similarity, which enables us to obtain a formula for $D^{(m)}$. It will be seen that both in $D^{(2)}$ and in $D^{(3)}$ some of the terms have sign + and some have sign −. We will see first how the terms are constructed, and then consider the question of sign.

I. 4. **Construction of terms.**—(i) For each of the three D's, for which the values of m are 1, 2, 3 respectively, each term is the product of m factors, which are the coefficients in the equations in § 2. In writing down these coefficients,

it is convenient to keep them in the relative positions in which they occur in the equations. Thus we get

For $D^{(1)}$	For $D^{(2)}$	For $D^{(3)}$
a_1	$a_1 \quad b_1$	$a_1 \quad b_1 \quad c_1$
	$a_2 \quad b_2$	$a_2 \quad b_2 \quad c_2$
		$a_3 \quad b_3 \quad c_3$

In each case we have a **set** of quantities arranged in the form of a square. The individual quantities are called the **elements** of the set; the quantities in a vertical line constitute a **column**, and the quantities in a horizontal line constitute a **row**. The columns are numbered from the left, and the rows from the top. The diagonal drawn from the top left-hand corner—i. e. the diagonal through a_1—is called the **leading diagonal.**

(ii) Each term contains m factors, which are taken from the set in such a way that one (only) shall come from each column and that one (only) shall come from each row. Also $D^{(m)}$ contains every term which can be constructed in this way. Take, for example, $D^{(3)}$. Since one factor is to come from each column, the factors are an a, a b, and a c. The a can be either a_1 or a_2 or a_3, i. e. it can be taken in three ways; when one of these three a's has been taken, one row has been used up, and the b can only be taken in two ways; and, when one of the two b's has been taken, the c can similarly only be taken in one way. There are therefore $3 . 2 . 1 = 6$ possible combinations of factors; and this is the number of terms in $D^{(3)}$.

(iii) Another way of stating the thing is that, if we keep to a fixed order $a\ b\ c$ of the factors in each term, the suffixes of the factors are the numbers 1 or 1 2 or 1 2 3, arranged in different ways, and there is one term for each of the possible arrangements.

(iv) We conclude that $D^{(m)}$ contains terms, each of which is constructed by taking m factors from the set of $m \times m = m^2$ quantities

$$\begin{array}{cccc} a_1 & b_1 & c_1 \ldots f_1 \\ a_2 & b_2 & c_2 \ldots f_2 \\ \vdots & \vdots & \vdots \quad\;\; \vdots \\ a_m & b_m & c_m \ldots f_m \end{array}$$

in such a way that there shall be one factor (only) taken from each column and one (only) from each row; there being one term for each of the $m\,(m-1)\ldots 1 \equiv m\,!$ ways in which this can be done. Or, which comes to the same thing, that the terms are made up of factors $a\,b\,c\ldots f$ with suffixes $1\;2\;3\ldots m$ arranged in different ways, there being one term for each of the $m\,!$ possible different arrangements.

I. 5. **Rule of signs.**—(i) It will be seen that, in the D's after $D^{(1)}$, half of the terms are positive and half negative, and that in each case the term found from the elements in the leading diagonal—namely a_1 or $a_1 b_2$ or $a_1 b_2 c_3$—is positive. We should therefore expect that half of the terms in $D^{(m)}$ would be positive and half negative, the term $a_1\,b_2\,c_3 \ldots f_m$—which we call the **leading term** and usually place first—being positive. The difficult question is that of signs. In the case of (say) $m = 5$, how are we to know whether such a term as $a_3 b_5 c_2 d_4 e_1$ is to have the sign $+$ or $-$?

(ii) The sign of a term must, if the letters $a\;b\;c\ldots f$ are kept in their original order, depend on the arrangement of the suffixes, i. e. on the extent to which this has departed from the initial arrangement $1\;2\;3\ldots m$. Now any arrangement such as 35241 can be got from the initial arrangement 12345 by a series of interchanges of adjacent figures. We must fix a definite order in which these interchanges are to

be made. We therefore say that each figure is to be moved in turn, beginning with that which ultimately comes first, then that which ultimately comes second, and so on. Thus in this particular case the successive stages would (repeating for each group of interchanges the arrangement from which we start) be 12345, 13245, 31245; 1245, 1254, 1524, 5124; 124, 214; 14, 41; 1: a total of seven interchanges. Now let us look at the signs in $D^{(2)}$ and $D^{(3)}$. In $D^{(2)}$ the arrangement 21 is obtained from 12 by 1 interchange, and the sign for 21 is $-$. In $D^{(3)}$ the signs of the successive terms, and the stages by which the final arrangements of suffixes are obtained, are as follows, the numbers of interchanges being added in heavy type:

$+$	123	**0**	$+$	123, 213, 231	**2**
$-$	123, 132 . .	**1**	$+$	123, 132, 312	**2**
$-$	123, 213 . .	**1**	$-$	123, 132, 312, 321 . .	**3**

It will be seen that both for $D^{(2)}$ and for $D^{(3)}$ the sign is $-$ or $+$ according as the number of interchanges is odd or even. We therefore adopt this as our rule; in the case of $a_3b_5c_2d_4e_1$, for instance, seven interchanges are necessary, and the sign is therefore $-$.

(iii) In order to find the sign of any given term by the above rule, it would be necessary to perform all the interchanges. A shorter method is to look at the term as it stands and to consider the *reversals of order* in it; i.e. taking the suffixes of the term in pairs in every possible way without altering their order, to see in how many cases the numbers are in the reverse of their order in the leading term, i.e. are in descending instead of ascending order. The term $a_3b_5c_2d_4e_1$, for instance, gives the following pairs, those in which the order is reversed being printed in heavier type:—35, **32**, 34, **31**, **52**, **54**, **51**, 24, **21**, **41**.

It is easily seen that each interchange, of the kind described in (ii) above, produces one reversal of order; for, while we are shifting one number, such as the 5 in the second group of arrangements there shown, the relative order of the other numbers remains unaltered. It follows that the number of reversals of order is the same as the number of interchanges of this kind; and therefore *the sign of a term will be $-$ or $+$ according as the number of reversals of order is odd or even.*

(iv) *The interchange of any two suffixes in a term changes the sign of the term.*

[Let the two suffixes be ϕ and ψ; ϕ coming before ψ in the term in question, but not necessarily being before it in numerical order.

(1) First let ϕ and ψ be adjacent. Then the interchange of ϕ and ψ increases or decreases the number of reversals of order by 1, and therefore changes the sign of the term.

(2) Next suppose that there are x suffixes between ϕ and ψ. Then we can move ψ in front of ϕ by $x + 1$ interchanges with the adjacent term, and then move ϕ into the original position of ψ by x interchanges. This is a total of $2x + 1$ interchanges, each of which in succession makes a change of sign: the total result is to change the sign of the term.]

(v) We have so far assumed that the factors of a term are arranged in the original order of the letters $a\ b\ c\ d \ldots$. Now suppose that the order of the factors is altered in any way. How does this affect the rule of signs?

The alteration of order can be brought about by a series of interchanges of factors. Suppose there is an interchange of a_ϕ and b_ψ. Then, by (iv), the number of reversals of order of suffixes is altered by an odd number, but the number of reversals of order of letters is also altered by an odd number; and therefore, if we consider the sum

of the numbers of reversals of order of letters and of suffixes, this sum either is not altered or is altered by an even number. It follows that, *if the factors of a term have been shifted about so that the letters $a\,b\,c\ldots$ are not in their original order, the sign of the term depends on the sum of the numbers of reversals of order of letters and of suffixes respectively*, being $-$ or $+$ according as this sum is odd or even. For example, in $d_4 b_5 c_2 a_3 e_1$ there are five reversals of order of the letters and eight of the suffixes, so that the sign is $-$.

II. PROPERTIES OF DETERMINANTS

II. 1. **Definition of determinant.**—We can now combine the results obtained in I. 4 and I. 5. We suppose that we are dealing with a set of $m \times m = m^2$ quantities, which we can arrange in the form of a square, thus (the quantities being denoted by crosses):

$$\begin{matrix} \times & \times & \times \dots \times \\ \times & \times & \times \dots \times \\ \vdots & \vdots & \vdots \quad\;\; \vdots \\ \times & \times & \times \dots \times \end{matrix}$$

Then the expression which we have to consider is the algebraical sum of a number of terms, of which some are taken positively and some negatively. Each term (apart from sign) is the product of m elements of the set, taken in such a way that one element (only) shall come from each column and that one element (only) shall come from each row; and there are $m!$ terms, corresponding to the $m!$ different ways in which this can be done. The leading term is the term containing the elements in the leading diagonal of the square, and has sign $+$. The signs of the other terms are to be found by replacing the elements of the set by $a_1, a_2, \dots b_1$, etc., arranged as a *key set*:

$$\begin{matrix} a_1 & b_1 & c_1 \dots f_1 \\ a_2 & b_2 & c_2 \dots f_2 \\ \vdots & \vdots & \vdots \quad\;\; \vdots \\ a_m & b_m & c_m \dots f_m \end{matrix}$$

The sign of a term is then $-$ or $+$ according as the sum of the numbers of reversals of order of the letters and of the suffixes, as compared with the leading term $a_1 b_2 c_3 \dots f_m$, is odd or even.

The algebraical sum of the terms so obtained, namely $a_1 b_2 c_3 \ldots f_m - a_2 b_1 c_3 \ldots f_m +$ etc., is called the **determinant** of the set, and will be denoted by D. It should however be observed that it is the determinant of the set *as so arranged*; with different arrangements of the elements of a set, still keeping them in a square, we may obtain different determinants.

We can therefore define the determinant as *the algebraical sum of terms of the form $a_p b_q c_r \ldots$, where $p\, q\, r \ldots$ are the numbers* 1 2 3 . . . m *arranged in some order, there being a term for each of the $m!$ possible orders, and the sign prefixed to the term being + for the natural order* 1 2 3 . . . m *and − or + for other orders according as the number of reversals of natural order is odd or even.*

The symbol for the determinant is constructed by placing single vertical lines before and after the set; thus

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

means the determinant $a_1 b_2 c_3 -$ etc., which we have called $D^{(3)}$.

The terminology is the same as is given in I. 4 and I. 5 for a set. The quantities between the vertical lines are the **elements** of the determinant. Those in a vertical line are a **column**; those in a horizontal line are a **row**. The **leading diagonal** is the diagonal drawn from the top left-hand corner; and the **leading term** is that containing the elements through which the leading diagonal passes. The leading term, as already stated, is taken positively.

If the symbol for a determinant contains m columns and m rows, the determinant is said to be of the mth **order**.

II. 2. Elementary properties.—(i) From the mode of

construction it follows that each element of the determinant appears in $(m-1)!$ out of the $m!$ terms; and there are no two terms having more than $m-2$ factors alike.

(ii) If each element of a column or of a row is 0, the determinant is $=0$. [For each term contains one of these elements as a factor.]

II. 3. Properties depending on the rule of signs.—

(i) *The value of a determinant is not altered by making the columns rows and the rows columns*; e.g., for $m=4$,

$$\begin{vmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{vmatrix}.$$

[Let D be the determinant, and R the new determinant obtained by making columns rows and rows columns in the symbol for D. Then, apart from sign, D and R obviously have the same $m!$ terms. We have therefore only to consider signs. The two determinants have the same leading term, which is positive in both. Let t be any other term of D, say $a_3b_4c_2d_1 \ldots$. Then t also occurs in R, but, since the terms of R must be constructed according to the system prescribed in our definition of a determinant, the factors of t in R will be arranged in the numerical order of the suffixes, namely $d_1c_2a_3b_4 \ldots$. The sign of t in D depends on the number of reversals of order in the suffixes 3 4 2 1 . . ., and the sign of t in R depends on the number of reversals of order in the letters $d\,c\,a\,b \ldots$. But each of these numbers is the sum of the numbers of reversals of order of letters and of suffixes as compared with the original orders $a\,b\,c\,d \ldots$ and 1 2 3 4 . . .; and, by I. 5 (v), these sums are either both odd or both even. It follows that the sign of t is the same in both determinants. This is true for each term of D or R; and the two determinants are therefore equal.]

If two determinants correspond so that the columns of one are the rows of the other, each determinant is said to be the **transposed** of the other.

(ii) It follows from (i) that any statement as to columns or rows is equally true for rows or columns. We shall indicate this, for conciseness, by 'column [row]' or 'row [column]'.

(iii) *If any two columns [rows] of a determinant are interchanged, the absolute magnitude of the determinant remains unaltered, but its sign is changed.*

[Suppose, e. g., that we interchange the b's and the e's. Let ϕ and ψ be any two suffixes. Then, in the original determinant, corresponding to any term which contains b_ϕ and e_ψ, there is another term exactly similar except that the factors are b_ψ and e_ϕ; and these two terms, by I. 5 (iv), are of opposite sign. The effect of interchanging the b's and the e's is that the two terms are interchanged, i. e. the sign of each is changed. This applies to every such pair of terms.]

(iv) *If any two columns [rows] of a determinant are identical, the determinant is* $= 0$.

[We can see this in either of two ways.

(1) Consider a pair of terms such as are mentioned in (iii). The one contains b_ϕ and e_ψ; the other is exactly similar, except that it contains b_ψ and e_ϕ; and the two terms have opposite signs. If $b_\phi = e_\phi$ and $b_\psi = e_\psi$, the two terms cancel. The whole determinant is made up of such pairs.

(2) More briefly, suppose we interchange the two columns which are identical. Then the determinant remains unaltered. But, by (iii), its sign is changed. This can only be the case if the determinant is 0.]

(v) Since columns and rows may be interchanged, a determinant is sometimes represented by its leading diagonal alone, if this indicates a system for insertion of the remaining elements. The notation is

$$|\, a_1 b_2 c_3 \dots f_m \,| \equiv \begin{vmatrix} a_1 & b_1 & c_1 & \dots f_1 \\ a_2 & b_2 & c_2 & \dots f_2 \\ a_3 & b_3 & c_3 & \dots f_3 \\ \vdots & \vdots & \vdots & \vdots \\ a_m & b_m & c_m & \dots f_m \end{vmatrix};$$

it is then immaterial, so far as the value of the determinant is concerned, whether we enter the a's as a column or as a row. But it should be mentioned that the relative arrangement of columns and rows is of importance later on, when we come to consider properties of sets of quantities.

II. 4. **Cofactors and minors.**—(i) In the complete expression for D, each term contains one a, which is either a_1 or a_2 or a_3 etc. We can therefore group the terms according to the a's they contain. In the case of $D^{(3)}$, for instance,

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1b_2c_3 - a_1b_3c_2 - a_2b_1c_3 + a_2b_3c_1 + a_3b_1c_2 - a_3b_2c_1$$

$$= a_1(b_2c_3 - b_3c_2) + a_2(-b_1c_3 + b_3c_1) + a_3(b_1c_2 - b_2c_1).$$

Suppose that the terms of D are grouped in this way; and let the resulting coefficients of $a_1\, a_2\, a_3 \ldots a_m$ be denoted by $A_1\, A_2\, A_3 \ldots A_m$. Then

$$D = a_1A_1 + a_2A_2 + a_3A_3 + \ldots + a_mA_m.$$

Similarly, if we group the terms according to the b's or c's etc. they contain and denote the coefficients of the b's or c's etc. by $B_1\, B_2\, B_3 \ldots B_m$ or $C_1\, C_2\, C_3 \ldots C_m$ etc., we shall have

$$D = b_1B_1 + b_2B_2 + b_3B_3 + \ldots + b_mB_m,$$
$$D = c_1C_1 + c_2C_2 + c_3C_3 + \ldots + c_mC_m,$$
etc.

The A's, B's, etc., are called the **cofactors** of the corresponding elements of the determinant; thus the cofactor of b_3 is B_3, where $b_3\, B_3$ is the sum of all the terms which contain b_3.

(ii) The terms which contain the leading element a_1 are obtained from the leading term $a_1b_2c_3 \ldots f_m$ by altering

the suffixes, and prefixing the proper sign to the term, in the manner already described, with the proviso that the factor a_1 remains unaltered. But this process will give us the products, by a_1, of the terms so constructed from a leading term $b_2 c_3 \ldots f_m$. In other words, the cofactor of a_1 is the determinant

$$\begin{vmatrix} b_2 & c_2 & \cdots f_2 \\ b_3 & c_3 & \cdots f_3 \\ \vdots & \vdots & \vdots \\ b_m & c_m & \cdots f_m \end{vmatrix}.$$

In $D^{(3)}$, for example, it is

$$b_2 c_3 - b_3 c_2 = \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix}.$$

(iii) The determinant which is obtained from D by striking out the column and the row which contain any element of the determinant is called the **minor** of that element in the determinant.

We see from (ii) that

$$A_1 \equiv \text{cofactor of } a_1 = \text{minor of } a_1.$$

We might show in the same way that the cofactor of any other element, say c_4, is equal to the minor of that element, with the sign $-$ or $+$ prefixed according to the position of the element in the determinant: but it is simpler to find the cofactor by bringing the element into the position of a_1. Let the element be in the qth column and the rth row. Then we can make it the leading element, without altering the order of the other columns or rows, by means of $q-1$ interchanges of its column with an adjoining column and $r-1$ interchanges of its row with an adjoining row. Each of these interchanges, by II. 3 (iii), multiplies the determinant by -1; and the total result is to multiply by -1 or by $+1$ according as $q+r-2$ is odd

or even. Having got the element into the position of the leading element, we strike out the first column and the first row; the result, apart from the prefixed sign, is still to give the minor of the element, since the relative positions of the other columns and rows are unaltered. Hence *the cofactor of any element is equal to its minor with the sign $-$ or $+$ prefixed according as the number of steps from the leading element to this element is odd or even*; it being understood that each step is either horizontally from one column to the next or vertically from one row to the next. For example,

$$A_3 = +\text{minor of } a_3,$$
$$C_4 = -\text{minor of } c_4,$$
etc.

(iv) We have found in (i) that

$$\left.\begin{array}{l} a_1A_1+a_2A_2+a_3A_3+\ldots+a_mA_m = D \\ b_1B_1+b_2B_2+b_3B_3+\ldots+b_mB_m = D \\ \text{etc.} \end{array}\right\}. \qquad \textbf{(II. 4. 1)}$$

We have now to find the value of such sums as

$$a_1B_1+a_2B_2+a_3B_3+\ldots+a_mB_m,$$
$$b_1A_1+b_2A_2+b_3A_3+\ldots+b_mA_m,$$
$$c_1A_1+c_2A_2+c_3A_3+\ldots+c_mA_m,$$
etc.

Let us take the second and third of these as examples, but replace the b's or the c's by θ's. Then we want to find the value of

$$\theta_1A_1+\theta_2A_2+\theta_3A_3+\ldots+\theta_mA_m.$$

Now we see from (II. 4. 1) that this is the value of the determinant

$$\begin{vmatrix} \theta_1 & b_1 & c_1 & \ldots f_1 \\ \theta_2 & b_2 & c_2 & \ldots f_2 \\ \theta_3 & b_3 & c_3 & \ldots f_3 \\ \vdots & \vdots & \vdots & \vdots \\ \theta_m & b_m & c_m & \ldots f_m \end{vmatrix};$$

for the cofactors of $\theta_1\,\theta_2\,\theta_3\,\ldots\,\theta_m$ in this determinant are the same as the cofactors of $a_1\,a_2\,a_3\,\ldots\,a_m$ in D, i.e. are $A_1\,A_2\,A_3\,\ldots\,A_m$. Let us replace θ throughout this determinant by any letter, other than a, occurring in D, e.g. by c. Then the determinant becomes

$$\begin{vmatrix} c_1 & b_1 & c_1 & \ldots f_1 \\ c_2 & b_2 & c_2 & \ldots f_2 \\ c_3 & b_3 & c_3 & \ldots f_3 \\ \vdots & \vdots & \vdots & \vdots \\ c_m & b_m & c_m & \ldots f_m \end{vmatrix}.$$

But this is a determinant which has two columns identical, and its value, by § 3 (iv), is 0. Hence

$$\left.\begin{array}{ll} & c_1 A_1 + c_2 A_2 + c_3 A_3 + \ldots + c_m A_m = 0 \\ \text{Similarly} & b_1 A_1 + b_2 A_2 + b_3 A_3 + \ldots + b_m A_m = 0 \\ & a_1 B_1 + a_2 B_2 + a_3 B_3 + \ldots + a_m B_m = 0 \\ & \text{etc.} \end{array}\right\}. \quad \textbf{(II. 4. 2)}$$

(v) Now let us interchange the columns with the rows, so that the determinant becomes

$$\begin{vmatrix} a_1 & a_2 & a_3 \ldots a_m \\ b_1 & b_2 & b_3 \ldots b_m \\ c_1 & c_2 & c_3 \ldots c_m \\ \vdots & \vdots & \vdots \quad \vdots \\ f_1 & f_2 & f_3 \ldots f_m \end{vmatrix}.$$

Then, by § 3 (i), the value of the new determinant is the same as that of the old, i.e. is D. Also the minor of any element ϵ in the new determinant is the same as its minor in the old determinant, but with columns and rows interchanged, so that its value is unaltered; and the number of steps from a_1 to ϵ is the same in both determinants. The cofactor of ϵ in the new determinant is therefore equal to its cofactor in the old determinant. Hence by applying

(II. 4. 1) and (II. 4. 2) to the new determinant we get new sets of relations, namely

$$\left.\begin{array}{l} a_1 A_1 + b_1 B_1 + c_1 C_1 + \ldots + f_1 F_1 = D \\ a_2 A_2 + b_2 B_2 + c_2 C_2 + \ldots + f_2 F_2 = D \\ \quad \text{etc.} \end{array}\right\} \qquad \text{(II. 4. 3)}$$

and

$$\left.\begin{array}{l} a_3 A_1 + b_3 B_1 + c_3 C_1 + \ldots + f_3 F_1 = 0 \\ a_2 A_1 + b_2 B_1 + c_2 C_1 + \ldots + f_2 F_1 = 0 \\ \quad \text{etc.} \end{array}\right\}. \qquad \text{(II. 4. 4)}$$

(vi) If all the elements in a column [row], except one, are 0, the determinant is equal to the product of that one by its cofactor.

III. SOLUTION OF SIMULTANEOUS EQUATIONS

III. 1. **Statement of previous results.**—We have next to consider the solution of the simultaneous equations (I. 3. A). Before we do this, it will be convenient to express in determinant form the results obtained in I. 2. These results are as follows:

(1) If $a_1x = k_1$, then* $x = |k_1| \div |a_1|$.

(2) If
$$\left.\begin{aligned} a_1x+b_1y &= k_1 \\ a_2x+b_2y &= k_2 \end{aligned}\right\},$$
then
$$x = \begin{vmatrix} k_1 & b_1 \\ k_2 & b_2 \end{vmatrix} \div \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}, \quad y = \begin{vmatrix} a_1 & k_1 \\ a_2 & k_2 \end{vmatrix} \div \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

(3) If
$$\left.\begin{aligned} a_1x+b_1y+c_1z &= k_1 \\ a_2x+b_2y+c_2z &= k_2 \\ a_3x+b_3y+c_3z &= k_3 \end{aligned}\right\},$$
then
$$x = \begin{vmatrix} k_1 & b_1 & c_1 \\ k_2 & b_2 & c_2 \\ k_3 & b_3 & c_3 \end{vmatrix} \div D^{(3)}, \quad y = \begin{vmatrix} a_1 & k_1 & c_1 \\ a_2 & k_2 & c_2 \\ a_3 & k_3 & c_3 \end{vmatrix} \div D^{(3)},$$
$$z = \begin{vmatrix} a_1 & b_1 & k_1 \\ a_2 & b_2 & k_2 \\ a_3 & b_3 & k_3 \end{vmatrix} \div D^{(3)},$$
where
$$D^{(3)} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

* Here, as elsewhere, vertical lines denote a determinant, not 'absolute value'.

III. 2. General solution.—The general equations of which we require a solution are those set out in (I. 3. A), namely:

$$\left.\begin{array}{l} a_1x + b_1y + c_1z + \dots + f_1w = k_1 \\ a_2x + b_2y + c_2z + \dots + f_2w = k_2 \\ \qquad\qquad\vdots \\ a_mx + b_my + c_mz + \dots + f_mw = k_m \end{array}\right\}. \qquad \textbf{(III. 2. A)}$$

The form of the solution is suggested by the results given above. Multiplying the successive equations by $A_1, A_2, \dots A_m$, and adding, we have

$$\begin{aligned} &(a_1A_1 + a_2A_2 + \dots + a_mA_m)x \\ &+ (b_1A_1 + b_2A_2 + \dots + b_mA_m)y \\ &+ (c_1A_1 + c_2A_2 + \dots + c_mA_m)z \\ &+ \dots \\ &+ (f_1A_1 + f_2A_2 + \dots + f_mA_m)w = k_1A_1 + k_2A_2 + \dots + k_mA_m. \end{aligned}$$

By (II. 4. 1) and (II. 4. 2) the coefficient of x is equal to D, and those of $y, z, \dots w$ are equal to 0. Also the expression on the right-hand side is what D would become if we replaced the a's by k's. Hence

$$\left.\begin{array}{l} x = (k_1A_1 + k_2A_2 + \dots + k_mA_m) \div D \\ \quad = \begin{vmatrix} k_1 & b_1 & c_1 & \dots f_1 \\ k_2 & b_2 & c_2 & \dots f_2 \\ \vdots & \vdots & \vdots & \vdots \\ k_m & b_m & c_m & \dots f_m \end{vmatrix} \div \begin{vmatrix} a_1 & b_1 & c_1 & \dots f_1 \\ a_2 & b_2 & c_2 & \dots f_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_m & b_m & c_m & \dots f_m \end{vmatrix}. \\ \text{Similarly} \\ y = \begin{vmatrix} a_1 & k_1 & c_1 & \dots f_1 \\ a_2 & k_2 & c_2 & \dots f_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_m & k_m & c_m & \dots f_m \end{vmatrix} \div \begin{vmatrix} a_1 & b_1 & c_1 & \dots f_1 \\ a_2 & b_2 & c_2 & \dots f_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_m & b_m & c_m & \dots f_m \end{vmatrix}, \\ \text{and so on.} \end{array}\right\} \qquad \textbf{(III. 2. 1)}$$

If, for verification, we substitute these values in the original equations, it will be found that the relations (II. 4. 3) and (II. 4. 4) come into play.

IV. PROPERTIES OF DETERMINANTS
(continued)

IV. 1. **Sum of determinants.**—If two determinants are identical except as regards one column [row], their sum is a similar determinant in which the elements of that column [row] are the sums of corresponding elements in the two determinants. [For example

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1+d_1 & b_1 & c_1 \\ a_2+d_2 & b_2 & c_2 \\ a_3+d_3 & b_3 & c_3 \end{vmatrix},$$

since it is $= (a_1A_1 + a_2A_2 + a_3A_3) + (d_1A_1 + d_2A_2 + d_3A_3)$
$= (a_1+d_1)A_1 + (a_2+d_2)A_2 + (a_3+d_3)A_3.$]

IV. 2. **Multiplication of determinant by a single factor.**—If each element of a column [row] is multiplied by the same factor, the determinant is multiplied by that factor. [For example

$$\begin{vmatrix} \lambda a_1 & b_1 & c_1 \\ \lambda a_2 & b_2 & c_2 \\ \lambda a_3 & b_3 & c_3 \end{vmatrix} = \lambda a_1A_1 + \lambda a_2A_2 + \lambda a_3A_3 = \lambda \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.]$$

IV. 3. **Alteration of column or row.**—If the elements of a column [row] are multiplied by a single factor and added to the corresponding elements of another column [row], the value of the determinant is not altered. [For example

$$\begin{vmatrix} a_1+\lambda c_1 & b_1 & c_1 \\ a_2+\lambda c_2 & b_2 & c_2 \\ a_3+\lambda c_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \lambda \begin{vmatrix} c_1 & b_1 & c_1 \\ c_2 & b_2 & c_2 \\ c_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.]$$

IV. 4. **Calculation of determinant.**—(i) There are two main methods for calculation of a numerical determinant.

(*a*) When m is small, we can use the formula

$$D = a_1 A_1 + a_2 A_2 + \dots + a_m A_m,$$

repeating the process as often as may be necessary. Thus, if

$$D \equiv \begin{vmatrix} 3 & -2 & 7 \\ 5 & 1 & -3 \\ 4 & 6 & 1 \end{vmatrix},$$

then $D = 3\begin{vmatrix} 1 & -3 \\ 6 & 1 \end{vmatrix} - 5\begin{vmatrix} -2 & 7 \\ 6 & 1 \end{vmatrix} + 4\begin{vmatrix} -2 & 7 \\ 1 & -3 \end{vmatrix}$

$$= 3 \,.\, 19 - 5(-44) + 4(-1) = 273.$$

(*b*) When m is large, we can reduce the determinant to one of order $m-1$ by means of § 3 and II. 4 (vi). Applying this method to the above example, we could multiply the first row by $\frac{5}{3}$ and subtract from the second, and also multiply it by $\frac{4}{3}$ and subtract from the third. To avoid fractions, we multiply D twice by 3. Then

$$9D = \begin{vmatrix} 3 & -2 & 7 \\ 15 & 3 & -9 \\ 12 & 18 & 3 \end{vmatrix} = \begin{vmatrix} 3 & -2 & 7 \\ 0 & 13 & -44 \\ 0 & 26 & -25 \end{vmatrix}$$

$$= 3 \times \begin{vmatrix} 13 & -44 \\ 26 & -25 \end{vmatrix} = 3(-325 + 1144) = 2{,}457.$$

$\therefore \quad D = 273.$

(ii) For algebraical determinants various devices have to be used. An important determinant is

$$D = \begin{vmatrix} a^{m-1} & b^{m-1} & c^{m-1} \dots e^{m-1} & f^{m-1} \\ a^{m-2} & b^{m-2} & c^{m-2} \dots e^{m-2} & f^{m-2} \\ \vdots & \vdots & \vdots \quad\; \vdots & \vdots \\ a & b & c \quad \dots e & f \\ 1 & 1 & 1 \quad \dots 1 & 1 \end{vmatrix}.$$

This is $= 0$ if $a = b$ or if $a = c$, etc. Hence it contains $a-b$, $a-c, \ldots a-e$, $a-f$ as factors. Similarly it contains $b-c, \ldots b-e$, $b-f$; and so on. Looking to the leading term, it will be seen that there can be no other factors; i.e.

$$D = (a-b)(a-c)\ldots(a-e)(a-f).(b-c)\ldots(b-e)(b-f)\ldots(e-f).$$

[Example.—Hence prove that

$$\begin{vmatrix} b^{m-1} & c^{m-1} & \ldots\ldots & e^{m-1} & f^{m-1} \\ b^{m-2} & c^{m-2} & \ldots\ldots & e^{m-2} & f^{m-2} \\ \vdots & \vdots & & \vdots & \vdots \\ b^{r+1} & c^{r+1} & \ldots\ldots & e^{r+1} & f^{r+1} \\ b^{r-1} & c^{r-1} & \ldots\ldots & e^{r-1} & f^{r-1} \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & \ldots\ldots & 1 & 1 \end{vmatrix} = (-)^{m-r+1}\frac{D}{\phi(a)} \times \text{coefficient of } a^r \text{ in } \phi(a),$$

where

$$\phi(a) \equiv (a-b)(a-c)\ldots(a-e)(a-f).]$$

IV. 5. Product of determinants.—(i) The product of a determinant of order m and a determinant of order n can be expressed as a determinant of order $m+n$ by placing the leading diagonals in line and filling in with 0's. For example

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \times \begin{vmatrix} d_4 & e_4 \\ d_5 & e_5 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 & 0 & 0 \\ a_2 & b_2 & c_2 & 0 & 0 \\ a_3 & b_3 & c_3 & 0 & 0 \\ 0 & 0 & 0 & d_4 & e_4 \\ 0 & 0 & 0 & d_5 & e_5 \end{vmatrix}.$$

[For the only terms of this latter determinant which are not 0 are those for which the first three factors (collectively) are taken from the first three columns and rows and the next two factors (collectively) from the last two columns and rows; and in each such case the first three factors form a term of the first determinant and the next two factors form a term of the second determinant. Thus all

the terms of the product of the one expanded determinant by the other are accounted for, and there are no others.]

(ii) We have now to show that the product of two determinants, each of order m, can be expressed as a determinant of order m. To obtain the general formula, it will be sufficient to take a particular case, e. g.

$$D \equiv \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}, \quad E \equiv \begin{vmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ \alpha_2 & \beta_2 & \gamma_2 \\ \alpha_3 & \beta_3 & \gamma_3 \end{vmatrix},$$

provided that in our reasoning we retain m as the order of each determinant.

By means of the first sentence of (i) we can write down DE as a determinant of order $2m$, i.e.

$$DE = \begin{vmatrix} a_1 & b_1 & c_1 & 0 & 0 & 0 \\ a_2 & b_2 & c_2 & 0 & 0 & 0 \\ a_3 & b_3 & c_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_1 & \beta_1 & \gamma_1 \\ 0 & 0 & 0 & \alpha_2 & \beta_2 & \gamma_2 \\ 0 & 0 & 0 & \alpha_3 & \beta_3 & \gamma_3 \end{vmatrix}.$$

Two of the quarters of this determinant contain 0's only; and it will be seen, from the method of forming those terms of the determinant which do not contain 0 as a factor, that we can fill in either of these quarters in any way we like, provided we leave the other quarter alone. Also, in order to reduce the determinant from order $2m$ to order m, we ought to get m 1's in the leading diagonal. We therefore shift the last m columns to be the first m, and replace the first m 0's in the new leading diagonal by 1's. The first process involves m^2 interchanges: we can avoid change of sign of the determinant, in the case where m is odd, by changing the signs of the first m rows (before inserting

the 1's), whether m is even or odd. Then, inserting the 1's, we get

$$DE = \begin{vmatrix} 1 & 0 & 0 & -a_1 & -b_1 & -c_1 \\ 0 & 1 & 0 & -a_2 & -b_2 & -c_2 \\ 0 & 0 & 1 & -a_3 & -b_3 & -c_3 \\ \alpha_1 & \beta_1 & \gamma_1 & 0 & 0 & 0 \\ \alpha_2 & \beta_2 & \gamma_2 & 0 & 0 & 0 \\ \alpha_3 & \beta_3 & \gamma_3 & 0 & 0 & 0 \end{vmatrix}.$$

We now reduce each of the elements in the right-hand top quarter to 0 by means of § 3; i.e. we add a_1 times the 1st column to the $(m+1)$th (in this case the 4th), thus getting

$$\begin{vmatrix} 1 & 0 & 0 & 0 & -b_1 & -c_1 \\ 0 & 1 & 0 & -a_2 & -b_2 & -c_2 \\ 0 & 0 & 1 & -a_3 & -b_3 & -c_3 \\ \alpha_1 & \beta_1 & \gamma_1 & a_1\alpha_1 & 0 & 0 \\ \alpha_2 & \beta_2 & \gamma_2 & a_1\alpha_2 & 0 & 0 \\ \alpha_3 & \beta_3 & \gamma_3 & a_1\alpha_3 & 0 & 0 \end{vmatrix},$$

then do the same with a_2 times the 2nd column and a_3 times the 3rd..., and then deal in the same way with the $(m+2)$th and $(m+3)$th ... columns. We get finally

$$DE = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \alpha_1 & \beta_1 & \gamma_1 & a_1\alpha_1+a_2\beta_1+a_3\gamma_1 & b_1\alpha_1+b_2\beta_1+b_3\gamma_1 & c_1\alpha_1+c_2\beta_1+c_3\gamma_1 \\ \alpha_2 & \beta_2 & \gamma_2 & a_1\alpha_2+a_2\beta_2+a_3\gamma_2 & b_1\alpha_2+b_2\beta_2+b_3\gamma_2 & c_1\alpha_2+c_2\beta_2+c_3\gamma_2 \\ \alpha_3 & \beta_3 & \gamma_3 & a_1\alpha_3+a_2\beta_3+a_3\gamma_3 & b_1\alpha_3+b_2\beta_3+b_3\gamma_3 & c_1\alpha_3+c_2\beta_3+c_3\gamma_3 \end{vmatrix}.$$

By the same reasoning as that employed at the beginning of this paragraph, we can replace each of the elements in the left-hand bottom quarter of the above determinant by 0. Hence, if the determinant formed by the elements

in the lower right-hand quarter of the above is called P, we have, by (i),

$$DE = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \times P = P;$$

i. e.

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \times \begin{vmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ \alpha_2 & \beta_2 & \gamma_2 \\ \alpha_3 & \beta_3 & \gamma_3 \end{vmatrix}$$
$$= \begin{vmatrix} a_1\alpha_1 + a_2\beta_1 + a_3\gamma_1 & b_1\alpha_1 + b_2\beta_1 + b_3\gamma_1 & c_1\alpha_1 + c_2\beta_1 + c_3\gamma_1 \\ a_1\alpha_2 + a_2\beta_2 + a_3\gamma_2 & b_1\alpha_2 + b_2\beta_2 + b_3\gamma_2 & c_1\alpha_2 + c_2\beta_2 + c_3\gamma_2 \\ a_1\alpha_3 + a_2\beta_3 + a_3\gamma_3 & b_1\alpha_3 + b_2\beta_3 + b_3\gamma_3 & c_1\alpha_3 + c_2\beta_3 + c_3\gamma_3 \end{vmatrix}. \quad \text{(IV.5.1)}$$

The reasoning is quite general, and the product of two determinants of any order can be written down from the above.

By interchanging columns and rows in one or other or both of the original determinants we get three other expressions. All four expressions, of course, are equal when expanded: we shall take the above to be the standard form. It is to be noticed that DE or $D \times E$ means the product of D and E, in this order; by reversing the order of multiplication we get four other forms, but these are only the 'transposed' of the previous four. The eight forms are given, in the new notation (see V. 6 (vi)), in the Appendix (p. 122). They are based on the principle that in the standard form the element in the qth column and rth row of the product is formed in a particular way from the qth column of the first determinant and the rth row of the second.

IV. 6. **The adjoint determinant.**—(i) Let D' denote the determinant whose elements are the cofactors of the corresponding elements in the original determinant, i. e.

$$D' \equiv \begin{vmatrix} A_1 & B_1 & C_1 \dots F_1 \\ A_2 & B_2 & C_2 \dots F_2 \\ \vdots & \vdots & \vdots \quad\; \vdots \\ A_m & B_m & C_m \dots F_m \end{vmatrix}. \qquad \text{(IV. 6. A)}$$

If we interchange columns and rows in D', and then express the product DD' as in (IV. 5. 1), we find that the elements in the principal diagonal of the product are $a_1A_1+a_2A_2+\dots a_mA_m$, $b_1B_1+b_2B_2+\dots+b_mB_m$, etc., each of which, by (II. 4. 1), is $=D$, while the other elements are $a_1B_1+a_2B_2+\dots+a_mB_m$, $b_1A_1+b_2A_2+\dots+b_mA_m$, etc., each of which, by (II. 4. 2), is $=0$. Hence

$$DD' = \begin{vmatrix} D & 0 & 0\dots 0 \\ 0 & D & 0\dots 0 \\ \vdots & \vdots & \vdots \quad \vdots \\ 0 & 0 & 0\dots D \end{vmatrix} = D^m, \qquad \text{(IV. 6. 1)}$$

and therefore

$$D' = D^{m-1}. \qquad \text{(IV. 6. 2)}$$

The determinant D' was formerly called the *reciprocal*, but is now more usually called the **adjoint**, of D. It is not the true reciprocal of D, since the product of the two is not 1 but D^m (cf. V. 2).

(ii) Let the cofactors of $A_1, B_1, C_1, \dots F_1$ in D' be denoted by $\alpha_1, \beta_1, \gamma_1, \dots \zeta_1$. Then, applying (II. 4. 3) and (II. 4. 4) to D', we have

$$\left.\begin{aligned} A_1\alpha_1+B_1\beta_1+C_1\gamma_1+\dots+F_1\zeta_1 &= D' \\ A_2\alpha_1+B_2\beta_1+C_2\gamma_1+\dots+F_2\zeta_1 &= 0 \\ &\;\,\vdots \\ A_m\alpha_1+B_m\beta_1+C_m\gamma_1+\dots+F_m\zeta_1 &= 0 \end{aligned}\right\}.$$

We can regard these as equations for determining $\alpha_1, \beta_1, \gamma_1, \dots \zeta_1$. Comparing them with

$$\left.\begin{aligned} A_1a_1+B_1b_1+C_1c_1+\dots+F_1f_1 &= D \\ A_2a_1+B_2b_1+C_2c_1+\dots+F_2f_1 &= 0 \\ &\;\,\vdots \\ A_ma_1+B_mb_1+C_mc_1+\dots+F_mf_1 &= 0 \end{aligned}\right\},$$

which are obtained from (II. 4. 3) and (II. 4. 4), we see that the solution is

$$\alpha_1 = \frac{D'}{D} a_1 = D^{m-2} a_1, \quad \beta_1 = \frac{D'}{D} b_1 = D^{m-2} b_1, \quad \text{etc.}$$

Hence *the cofactor of any element of the adjoint determinant is equal to the corresponding element of the original determinant, multiplied by the ratio of the adjoint determinant to the original determinant.*

V. THE TENSOR NOTATION *

V. 1. **Main properties of determinant.**—(i) The notation so far used is the ordinary one for elementary work. For higher work we reduce the number of letters and make a more liberal use of suffixes. We shall replace $x, y, z \ldots w$ by $X_1, X_2, X_3, \ldots X_m$; $a_1, b_1, c_1, \ldots f_1$ by $d_{11}, d_{21}, d_{31}, \ldots d_{m1}$; and so on. Also, it being understood that the values assignable to each of the letters q and r are 1, 2, 3, ... m, we can use $|d_{qr}|$ and $|d_{rq}|$ to denote the determinants whose elements in the qth column and rth row are respectively † d_{qr} and d_{rq}, i.e.

$$\left.\begin{aligned} |d_{qr}| &\equiv \begin{vmatrix} d_{11} & d_{21} & d_{31} \ldots d_{m1} \\ d_{12} & d_{22} & d_{32} \ldots d_{m2} \\ \vdots & \vdots & \vdots \quad\;\; \vdots \\ d_{1m} & d_{2m} & d_{3m} \ldots d_{mm} \end{vmatrix} \\ |d_{rq}| &\equiv \begin{vmatrix} d_{11} & d_{12} & d_{13} \ldots d_{1m} \\ d_{21} & d_{22} & d_{23} \ldots d_{2m} \\ \vdots & \vdots & \vdots \quad\;\; \vdots \\ d_{m1} & d_{m2} & d_{m3} \ldots d_{mm} \end{vmatrix} \end{aligned}\right\}, \qquad \text{(V. 1. A)}$$

so that the interchangeability of columns and rows gives

$$|d_{qr}| = |d_{rq}|. \qquad \text{(V. 1. 1)}$$

(ii) The k's in (III. 2. A) were supposed to be known

* The paragraphs in small print may be omitted on first reading, but should be read before Chapter VI is taken.

† It is more usual to have the suffixes the other way round; i.e. to use 11, 12, 13, ... $1m$ as suffixes for the first row. But my arrangement seems to follow more naturally from the $a_1 b_1 \ldots$ notation, and I also find that it fits in better with the subsequent work.

quantities, so that the equations served to determine $x, y, z, \dots w$. We shall have to consider the relations between the set of quantities which we have denoted by $x, y, z, \dots w$ and those which we have denoted by $k_1, k_2, k_3, \dots k_m$. We therefore, in altering $x, y, z, \dots w$ to $X_1, X_2, X_3, \dots X_m$, also alter $k_1, k_2, k_3, \dots k_m$ to $Y_1, Y_2, Y_3, \dots Y_m$.

(iii) The main properties which we have to consider are set out below. Definitions are marked with capital letters: propositions with arabic numbers.

$$D \equiv | d_{qr} | \equiv \begin{vmatrix} d_{11} & d_{21} & d_{31} \dots d_{m1} \\ d_{12} & d_{22} & d_{32} \dots d_{m2} \\ \vdots & \vdots & \vdots \\ d_{1m} & d_{2m} & d_{3m} \dots d_{mm} \end{vmatrix} \quad . \quad \text{(A)}$$

$$D_{ps} \equiv \text{cofactor of } d_{ps} \text{ in } D \quad . \quad . \quad . \quad \text{(B)}$$

$$\begin{pmatrix} p = 1, 2, \dots m : \\ q = 1, 2, \dots m \end{pmatrix} D_{p1}d_{q1} + D_{p2}d_{q2} + \dots + D_{pm}d_{qm}$$

$$= \begin{cases} D \text{ if } q = p \\ 0 \text{ if } q \neq p \end{cases} \quad \text{(1)}$$

$$D' \equiv | D_{qr} | \quad . \quad . \quad . \quad . \quad . \quad \text{(C)}$$

$$DD' = D^m \quad . \quad . \quad . \quad . \quad . \quad \text{(2)}$$

$$\text{cofactor of } D_{ps} \text{ in } D' = D^{m-2}d_{ps} \quad . \quad . \quad \text{(3)}$$

If
$$\left.\begin{aligned} d_{11}X_1 + d_{21}X_2 + d_{31}X_3 + \dots + d_{m1}X_m &= Y_1 \\ d_{12}X_1 + d_{22}X_2 + d_{32}X_3 + \dots + d_{m2}X_m &= Y_2 \\ &\vdots \\ d_{1m}X_1 + d_{2m}X_2 + d_{3m}X_3 + \dots + d_{mm}X_m &= Y_m \end{aligned}\right\},$$

then

$$(p = 1, 2, \dots m)$$

$$X_p = \frac{D_{p1}Y_1 + D_{p2}Y_2 + D_{p3}Y_3 + \dots + D_{pm}Y_m}{D} \quad . \quad \text{(4)}$$

(iv) As a preliminary, to make the statements more concise, we may—though this is not essential—introduce

the ordinary notation of summation. Thus (1) can be written

$$\begin{pmatrix} p = 1, 2, \ldots m; \\ q = 1, 2, \ldots m \end{pmatrix} \sum_{s=1}^{m} D_{ps} d_{qs} = \begin{Bmatrix} D \text{ if } q = p \\ 0 \text{ if } q \neq p \end{Bmatrix}; \quad . \quad (1)$$

and (4) can be written

$$\left.\begin{aligned} &\text{If } (s = 1, 2, \ldots m) \sum_{p=1}^{m} d_{ps} X_p = Y_s, \\ &\text{then } (p = 1, 2, \ldots m) X_p = \sum_{s=1}^{m} \frac{D_{ps} Y_s}{D}. \end{aligned}\right\} \quad . \quad . \quad (4)$$

(v) The tensor notation involves five steps, which are set out in §§ 2–4, 6, 8 below. The reader will find it helpful to copy out the statement in (iii), modified as in (iv), and make the successive alterations which are now to be described.

V. 2. Reciprocal determinant.—The first step (which will be found in modern text-books) relates to (1) and (2). The determinant D' has sometimes been called the reciprocal of D. But, as has already been pointed out in IV. 6 (i), it is not a true reciprocal, since the product of the two determinants is not 1 but D^m. Since, however, D' contains m columns, we see that if we form a new determinant by dividing each element of it by D the product of this new determinant and D will be 1. We therefore write (1) in the form

$$\begin{pmatrix} p = 1, 2, \ldots m; \\ q = 1, 2, \ldots m \end{pmatrix} \sum_{s=1}^{m} \frac{D_{ps}}{D} d_{qs} = \begin{Bmatrix} 1 \text{ if } q = p \\ 0 \text{ if } q \neq p \end{Bmatrix}; \quad . \quad (1)$$

and form a new determinant D'' defined by

$$D'' \equiv | D_{qr}/D |. \quad . \quad . \quad . \quad . \quad . \quad (C)$$

This new determinant will be equal to $D'/D^m = 1/D$, so

that the product of the two determinants will be **1**; and the cofactor of D_{ps}/D in the new determinant will be $D^{m-2}d_{ps}/D^{m-1} = d_{ps}/D = d_{ps}D''$. Hence D and D'' are so related that their product is 1 and that the cofactor of any element of either determinant, divided by the determinant, is equal to the corresponding element of the other determinant. We can therefore call each determinant the **reciprocal** of the other.

V. 3. Elements of reciprocal determinant.—The next step is to have a single symbol for

$$(\text{cofactor of } d_{ps} \text{ in } D) \div D.$$

We have already used D_{ps} for the cofactor of d_{ps}; and it is inconvenient to introduce a new letter in place of d or D. We therefore denote the above expression by

$$d^{ps}.$$

We accordingly, in our statement, replace (B), (1), (C), (2), (3), and the second line of (4), by

$$d^{ps} \equiv (\text{cofactor of } d_{ps} \text{ in } D) \div D \quad . \quad . \quad . \quad \text{(B)}$$

$$\begin{pmatrix} p = 1, 2, \ldots m; \\ q = 1, 2, \ldots m \end{pmatrix} \sum_{s=1}^{m} d^{ps} d_{qs} = \begin{Bmatrix} 1 \text{ if } q = p \\ 0 \text{ if } q \neq p \end{Bmatrix} \quad . \quad . \quad (1)$$

$$D'' \equiv | d^{qr} | \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(C)}$$

$$DD'' = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$d_{ps} = (\text{cofactor of } d^{ps} \text{ in } D'') \div D'' \quad . \quad . \quad (3)$$

$$(p = 1, 2, \ldots m)\, X_p = \sum_{s=1}^{m} d^{ps} Y_s \quad . \quad [\text{2nd line of}] \quad (4)$$

The parallelism of (B) and (3), and of the two lines of (4), should be noted.

V. 4. Set-notation.—(i) The next step is to abbreviate (4), as altered. This contains two statements—a hypothesis

and a conclusion; they are similar in form, so that we need only consider the first one, namely

$$(s = 1, 2, \dots m) \sum_{p=1}^{m} d_{ps} X_p = Y_s.$$

(ii) The expression on the left-hand side of this statement has a definite value for each value of s; we can denote these values by

$$E_1\ E_2\ E_3 \dots E_m.$$

The statement then takes the form

$$(s = 1, 2, \dots m) E_s = Y_s.$$

We cannot merely omit the '$(s = 1, 2, \dots m)$' from this, without leaving it doubtful whether we are speaking of some particular s or of each s. We get over the difficulty by omitting the '$(s = 1, 2, \dots m)$' and replacing the s in '$E_s = Y_s$' by a *Greek* letter. The convention then is that a statement of the form

$$E_\sigma = Y_\sigma$$

means that E_s is equal to Y_s for each of the values of s, i.e. that

$$E_1 = Y_1,\ E_2 = Y_2,\ E_3 = Y_3, \dots E_m = Y_m;$$

it being understood that the values $1, 2, 3, \dots m$ which are to be given to a Greek letter have been settled beforehand and remain the same throughout our work.

(iii) Applying this convention to the two statements in (4), it becomes:

$$\text{If } \sum_{p=1}^{m} d_{p\sigma} X_p = Y_\sigma, \text{ then } X_\lambda = \sum_{s=1}^{m} d^{\lambda s} Y_s \quad . \quad . \quad . \quad (4)$$

It is immaterial what Greek letter we use in either of these statements, provided the letter is the same on both sides. We could have used the same letter in the two

statements, but in the particular case it is better to have the letters different.*

We have rather spoilt the symmetry of (4), but we will put this right in § 6.

(iv) The statement in (1) is a statement as to the value of a certain expression for all values of p and all values of q; and, so far as the left-hand side is concerned, we could extend the above principle by using two Greek letters. But the right-hand side presents difficulties; and we must therefore leave this over for later consideration (§ 8).

V. 5. **Principles of set-notation.**—It is desirable at this stage to consider the principles underlying the notation which we have just adopted.

(i) Take first the case of a **single set** of m quantities or **elements**; i.e. an aggregate of m quantities which fall into a certain linear arrangement. We denote these by, say,

$$A_1 \; A_2 \; A_3 \ldots A_m.$$

We have settled that a statement such as

$$A_\lambda = E_\lambda$$

is a comprehensive way of saying that

$$A_1 = E_1,\; A_2 = E_2,\; A_3 = E_3, \ldots A_m = E_m.$$

Thus we use A_p etc. when we are referring to a particular member of the set, and we use A_λ etc. when we are making a statement with regard to each member of the set in turn. We may also want to speak of the set as a whole. It will be found that no confusion arises from using A_λ in this sense also. We can therefore say that

$$A_\lambda \equiv (A_1 \; A_2 \; A_3 \ldots A_m);$$

the brackets being used in order to show that we are considering the set as a whole. In this sense we might regard the statement $A_\lambda = E_\lambda$ as meaning that the set A_λ as a whole is equal to the set

* I have as far as possible used λ, μ, ρ, σ in this chapter to correspond to p, q, r, s, reserving ν for product-sums (§ 6). Later on it is better to have no fixed rule, beyond that laid down in § 5 (iii).

E_λ as a whole, this equality of the wholes implying equality of the parts. We shall have to take this step later (Chapter VI): for the present it will be sufficient to regard the statement $A_\lambda = E_\lambda$ as merely an abbreviated way of saying that $A_1 = E_1$, $A_2 = E_2$, etc.

(ii) The interpretation of A_λ as the set as a whole recalls the idea of a vector as the resultant of a number of components. But the operations which we shall have to perform with single sets do not follow exactly the same laws as those which govern operations with vectors as ordinarily understood (see note to (vi) below), so that the analogy must not be pushed too far.

(iii) Next take the case of a **double set** of m^2 quantities, i.e. a set consisting of m single sets, each containing m elements. If we denote the elements of the qth single set by $F_{q1}, F_{q2}, \ldots F_{qm}$, this single set as a whole can be called $F_{q\rho}$, and the complete double set can be called $F_{\mu\rho}$. We think of the elements as arranged in a square, the columns of which are the single sets: by regrouping (e.g. $F_{1r}, F_{2r}, \ldots F_{mr}$) we get the rows of the square. We have already, in § 1, adopted the convention that in d_{qr} or d_{rq} the q represents the column and the r the row; and similarly we shall say that in $F_{\mu\rho}$ or $F_{\rho\mu}$ etc. the first letter, according to alphabetical order, means the column and the second the row. Hence

$$F_{\mu\rho} \equiv \begin{pmatrix} F_{11} & F_{21} & F_{31} & \ldots F_{m1} \\ F_{12} & F_{22} & F_{32} & \ldots F_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ F_{1m} & F_{2m} & F_{3m} & \ldots F_{mm} \end{pmatrix}, \quad F_{\rho\mu} \equiv \begin{pmatrix} F_{11} & F_{12} & F_{13} & \ldots F_{1m} \\ F_{21} & F_{22} & F_{23} & \ldots F_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ F_{m1} & F_{m2} & F_{m3} & \ldots F_{mm} \end{pmatrix},$$

the brackets being inserted, as before, in order to show that the set of quantities is in each case to be regarded as a whole. Then the statements

$$F_{\mu\rho} = G_{\mu\rho}, \quad F_{\mu\rho} = H_{\rho\mu},$$

mean respectively that $F_{qr} = G_{qr}$, and that $F_{qr} = H_{rq}$, for every value of q taken with every value of r.

A double set is **symmetrical** if it is not altered by interchanging columns and rows.

(iv) A particular form of double set is obtained by multiplying together every element of one single set and every element of another single set (of the same number of elements). If these two sets are B_μ and C_ρ (in this order), the representative element of

the resulting double set will be $B_q C_r$, so that the double set can be represented by $B_\mu C_\rho$ or $C_\rho B_\mu$. This double set is called the **product** of the two single sets. It should be noted that we must not write it as '$B_\mu C_\mu$' or as '$B_\rho C_\rho$'; partly because this would not define the particular arrangement of the elements of the double set, and partly because we shall presently have to give a special meaning to these latter expressions.

(v) In addition to double sets and single sets, we have to use single quantities, such as a or k. Any such quantity is called a **scalar**. It need not be a constant: it may, as will be seen later, be a definite function of the elements of one or more sets.

(vi) We shall for the present be dealing only with expressions which, interpreted according to the laws of ordinary algebra, are obtained from scalars, single sets, and double sets by addition, subtraction, and multiplication. The rule of interpretation is the same that we adopted in (iv) for $B_\mu C_\rho$: we replace the Greek letters $\lambda\ \mu\ \rho \ldots$ by $p\ q\ r \ldots$ and take the total expression to be the set obtained by giving to each of the quantities $p\ q\ r \ldots$ each of the values $1\ 2\ 3 \ldots m$. For example:

(*a*) kA_ρ means the single set whose elements are

$$kA_1,\ kA_2, \ldots kA_m;$$

(*b*) $A_\lambda \pm bB_\lambda$ means the single set whose elements are *

$$A_1 \pm bB_1,\ A_2 \pm bB_2, \ldots A_m \pm bB_m;$$

(*c*) $A_{\mu\rho} - aF_\rho G_\mu$ means the double set whose element in the qth column and the rth row is $A_{qr} - aF_r G_q$.

It is obvious that this system of interpretation is in accordance with the laws of ordinary algebra; for instance

$$k(A_\mu + B_\mu) = kA_\mu + kB_\mu = kB_\mu + kA_\mu,$$
$$F_\lambda(G_\rho - H_\rho) = F_\lambda G_\rho - F_\lambda H_\rho,$$

and so on.

We are further restricted, in the case of expressions containing

* It will be seen from (vi) (*a*) and (*b*) and from (iv) that single sets follow the same rule as ordinary vectors as regards multiplication by a scalar, addition, and subtraction, but not as regards multiplication together.

more than one term, to (1) scalar expressions, (2) single sets arising as sums ('sum' of course including 'difference') of expressions which contain the same letter, e.g. $aA_\lambda + bB_\lambda + \ldots$, (3) double sets arising as sums of expressions which contain the same pair of letters, e.g. $aA_{\mu\nu} + bB_\nu C_\mu$. We do not therefore have to consider such an expression as $A_\lambda + B_\sigma$, which is really a double set, not a single set.

(vii) The suffixes which we have so far attached to a symbol have usually, in accordance with the regular practice in algebra and with the ordinary meaning of the word, been placed below the line: the exception being the use of d^{ps} to mean (cofactor of d_{ps} in D) $\div D$. This latter system of having upper suffixes as well as lower suffixes will sometimes be found convenient. We may, for instance, want to denote a single set by A^λ; and in that case $A^1, A^2, \ldots$ would be members of the set. Where there is any risk of confusion, we shall not use the ordinary indices of algebra at all; thus the square of A_p will be $A_p A_p$, not $A_p{}^2$.

V. 6. **Product-sum notation.**—(i) Our next simplification consists in dropping the sign of summation in (1) and (4). But, since merely to drop it and to replace, say, $\sum_{s=1}^{m} d^{ps}d_{qs}$ by $d^{ps}d_{qs}$ would be misleading, we use a special notation. The number of alphabets at our disposal is limited: and it will be found not only that we can use Greek letters for this purpose without risk of error, but that there are actual advantages in doing so.

(ii) The rule we adopt is that, when an expression of the form B_pC_p has to be summed for the values $1, 2, \ldots m$ of p, we denote the result by replacing p by a Greek letter in both places; and, conversely, the meaning of such an expression as $B_\nu C_\nu$ is

$$B_\nu C_\nu \equiv B_1C_1 + B_2C_2 + \ldots + B_mC_m. \qquad \textbf{(V. 6. A)}$$

(iii) The pair of ν's in $B_\nu C_\nu$ could be replaced by a pair of any other identical Greek letters; e.g.

$$B_\nu C_\nu = B_\rho C_\rho. \qquad \textbf{(V. 6. 1)}$$

The ν (or ρ) is for this reason called a **dummy**. We can think of the sum represented by $B_\nu C_\nu$ as the result of linking the elements of B_ν with the corresponding elements of C_ν; we can therefore describe a Greek letter which occurs twice as a **linked suffix**, and one which occurs once only as a **free suffix.**

(iv) The rule in (ii) applies if either or both of the expressions B_p and C_p has a free suffix as well as the p; e.g. $A_\nu B_{\mu\nu}$ means $A_1 B_{\mu 1} + A_2 B_{\mu 2} + \dots + A_m B_{\mu m}$, and $A_{\lambda\nu} B_{\rho\nu}$ means $A_{\lambda 1} B_{\rho 1} + A_{\lambda 2} B_{\rho 2} + \dots + A_{\lambda m} B_{\rho m}$, which is a double set whose typical element is $A_{q\nu} B_{r\nu}$.

(v) We can also have successive summations expressed in the same way. Thus

$$B_\lambda C_{\lambda\mu} D_{\mu\rho} E_{\sigma\rho}$$

involves summations with regard to λ, with regard to μ, and with regard to ρ. It is easy to show that these summations can be made in any order: e.g. we can take $C_{\lambda\mu}$ and $D_{\mu\rho}$ together as if the λ and ρ were free, and then bring in B_λ and $E_{\sigma\rho}$.

(vi) As an example of the brevity effected by this notation we may take the expression for the product of determinants. Even for so small a value of m as 3, the expression obtained in IV. 5 (ii) for the product of two determinants is formidable. We can condense it by introducing Σ's, but the result is clumsy. In the new notation it will be found that the method of IV. 5 (ii) gives

$$| a_{qr} | \times | b_{qr} | = | a_{q\lambda} b_{\lambda r} |. \qquad \textbf{(V. 6. 2)}$$

The process can be repeated: e.g.

$$|\,a_{qr}\,|\times|\,b_{qr}\,|\times|\,c_{qr}\,|\times|\,d_{qr}\,| = |\,a_{q\lambda}b_{\lambda\mu}c_{\mu\nu}d_{\nu r}\,|. \quad \textbf{(V. 6. 3)}$$

(vii) The result of applying the product-sum notation to the statement in § 1 (iii) is that (1) and (4) become (p and s in (4) being replaced by λ and σ respectively)

$$\begin{pmatrix} p = 1, 2, \dots m; \\ q = 1, 2, \dots m \end{pmatrix} d^{p\sigma}d_{q\sigma} = \begin{Bmatrix} 1 \text{ if } q = p \\ 0 \text{ if } q \neq p \end{Bmatrix}. \quad (1)$$

$$\text{If } d_{\lambda\sigma}X_\lambda = Y_\sigma, \text{ then } X_\lambda = d^{\lambda\sigma}Y_\sigma. \quad (4)$$

V. 7. **Inner products of sets.**—(i) The quantity $B_\nu C_\nu$ behaves in many respects like an algebraical product. We call it the **inner product** of B_ν and C_ν, to distinguish it from an ordinary or **outer** product such as $B_\lambda C_\rho$ (§ 5 (iv)). The inner product of B_ν and C_ν is the sum of the elements in the leading diagonal of the outer product of B_λ and C_ρ.

(ii) In the same way $A_\nu B_{\mu\nu}$ is the inner product of A_ν and $B_{\mu\nu}$, and $A_{\lambda\nu}B_{\nu\rho}$ is the inner product of $A_{\lambda\nu}$ and $B_{\nu\rho}$.

(iii) The process of forming an inner product, as above, may be called **inner multiplication.**

V. 8. Unit-set notation.—(i) We have finally to consider the form of (1), which is a statement that

$$\begin{pmatrix} p = 1, 2, \dots m; \\ q = 1, 2, \dots m \end{pmatrix} d^{p\sigma}d_{q\sigma} = \begin{Bmatrix} 1 \text{ if } q = p \\ 0 \text{ if } q \neq p \end{Bmatrix}.$$

So far as the left-hand side is concerned, this is a statement as to the values of the elements of a double set

$$d^{\lambda\sigma}d_{\mu\sigma}.$$

As regards the right-hand side, however, the statement falls into two; first that $d^{p\sigma}d_{p\sigma} = 1$, and next that $d^{p\sigma}d_{q\sigma} = 0$ if p and q are different. We want to replace these by a single statement.

(ii) We do this by converting the statement into one as

to the equality of two double sets. For this purpose we construct a set whose typical element, in the qth column and rth row, is 1 if q and r are the same and 0 if they are different; a set, in other words, which has 1 for each element of its leading diagonal and 0 everywhere else. If we call this set *

$$|^{\lambda}_{\mu},$$

then our definition of $|^{p}_{q}$ is that

$$|^{p}_{q} \equiv \text{the function of } p \text{ and } q \text{ which}$$
$$\text{is} = \begin{Bmatrix} 1 \text{ if } q = p \\ 0 \text{ if } q \neq p \end{Bmatrix}. \qquad \textbf{(V. 8. A)}$$

We can therefore write (1) in the form

$$\begin{pmatrix} p = 1,\ 2, \dots m; \\ q = 1,\ 2, \dots m \end{pmatrix} d^{p\sigma} d_{q\sigma} = |^{p}_{q};$$

or, in the set-notation,

$$d^{\lambda\sigma} d_{\mu\sigma} = |^{\lambda}_{\mu}. \qquad \dots\dots \qquad (1)$$

V. 9. **Properties of the unit set.**—(i) We have defined $|^{\lambda}_{\mu}$ as the set whose typical element is

$$|^{p}_{q} \equiv \begin{Bmatrix} 1 \text{ if } q = p \\ 0 \text{ if } q \neq p \end{Bmatrix}. \qquad \textbf{(V. 9. A)}$$

Hence each element of the leading diagonal of $|^{\lambda}_{\mu}$ is 1, and the other elements are all 0; in other words

$$|^{\lambda}_{\mu} \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \dots 0 \\ 0 & 1 & 0 & 0 \dots 0 \\ 0 & 0 & 1 & 0 \dots 0 \\ \vdots & \vdots & \vdots & \vdots \quad \vdots \\ 0 & 0 & 0 & 0 \dots 1 \end{pmatrix}. \qquad \textbf{(V. 9. B)}$$

* The usual symbol, adopted by Einstein, is δ^{λ}_{μ}; J. E. Wright ('Invariants of quadratic differential forms') uses $\eta_{\lambda\mu}$. Neither of these seems sufficiently distinctive; and δ already has a considerable number of other uses. I have therefore altered the symbol to $|^{\lambda}_{\mu}$ ('unit $\lambda\mu$'), as an experiment.

This set (with any pair of letters) will be called the **unit set.** The following are its chief properties.

(ii) The set is symmetrical, i. e.

$$|^{\lambda}_{\mu} = |^{\mu}_{\lambda}. \qquad \text{(V. 9. 1)}$$

(iii) The determinant of the set is

$$\left| |^{q}_{r} \right| = \begin{vmatrix} 1 & 0 & 0 \dots 0 \\ 0 & 1 & 0 \dots 0 \\ 0 & 0 & 1 \dots 0 \\ \vdots & \vdots & \vdots \quad \vdots \\ 0 & 0 & 0 \dots 1 \end{vmatrix} = 1. \qquad \text{(V. 9. 2)}$$

(iv) Also, if we multiply together this determinant and any other determinant, in either order, it will be found that we merely reproduce the latter, i. e.

$$\left| |^{q}_{r} \right| \times |a_{qr}| = |a_{qr}| \times \left| |^{q}_{r} \right| = |a_{qr}|. \qquad \text{(V. 9. 3)}$$

(v) The special importance of the set, or of any column or row of it, lies in its effect when combined with another set to form a product-sum. It will be found that, t having any one of the values 1, 2, 3, ... m,

$$|^{t}_{\mu} A_{\mu} = |^{\mu}_{t} A_{\mu} = A_t, \quad |^{t}_{\mu} A_{\mu\nu} = |^{\mu}_{t} A_{\mu\nu} = A_{t\nu}, \quad \text{(V. 9. 4)}$$

$$|^{\lambda}_{\mu} A_{\mu} = |^{\mu}_{\lambda} A_{\mu} = A_{\lambda}, \quad |^{\lambda}_{\mu} A_{\mu\nu} = |^{\mu}_{\lambda} A_{\mu\nu} = A_{\lambda\nu}. \quad \text{(V. 9. 5)}$$

[For example, take $t = 3$ in the first part of (V. 9. 4). Then

$$\begin{aligned} |^{3}_{\mu} A_{\mu} &= |^{3}_{1} A_1 + |^{3}_{2} A_2 + |^{3}_{3} A_3 + |^{3}_{4} A_4 + \dots \\ &= 0 . A_1 + 0 . A_2 + 1 . A_3 + 0 . A_4 + \dots \\ &= A_3 .] \end{aligned}$$

Thus the effect of inner multiplication by $|^{\lambda}_{\mu}$ of a single or double set which contains μ (or λ) is to alter the latter to λ (or μ).

V. 10. Determinant properties.—(i) Before we write down the final results, there is another small change which we shall find it convenient to make. In the statement $d^{\lambda\sigma} d_{\mu\sigma} = |^{\lambda}_{\mu}$, obtained in § 8 (ii), the linked suffixes are an upper σ and a lower σ, which cancel one another; and the

free suffixes λ and μ are in the same respective positions on the two sides. It is desirable that, whenever possible, these two conditions should exist. In each of the statements $d_{\lambda\sigma} X_\lambda = Y_\sigma$ and $X_\lambda = d^{\lambda\sigma} Y_\sigma$, given at the end of § 6, one of the conditions exists but the other does not. We make them both exist by replacing X_λ, throughout, by

$$X^\lambda \equiv (X^1 X^2 \ldots X^m),$$

as explained in § 5 (vii).

(ii) Our statement, after carrying out the alterations indicated in §§ 2–4, 6 and 8, and in (i) above, becomes—

Notation.	$D \equiv \lvert d_{qr} \rvert.$	(V. 10. A)
	$d^{ps} \equiv (\text{cofactor of } d_{ps} \text{ in } D) \div D.$	(V. 10. B)
	$D'' \equiv \lvert d^{qr} \rvert.$	(V. 10. C)
Properties.	$d^{\lambda\sigma} d_{\mu\sigma} = 1^\lambda_\mu.$	(V. 10. 1)
	$DD'' = 1.$	(V. 10. 2)
	$d_{ps} = (\text{cofactor of } d^{ps} \text{ in } D'') \div D''.$	(V. 10. 3)
	If $Y_\sigma = d_{\lambda\sigma} X^\lambda$, then $X^\lambda = d^{\lambda\sigma} Y_\sigma$.	(V. 10. 4)

To these we may add the formula (V. 6. 2) for multiplication of determinants, namely

$$\lvert a_{qr} \rvert \times \lvert b_{qr} \rvert = \lvert a_{q\lambda} b_{\lambda r} \rvert. \qquad \text{(V. 10. 5)}$$

V. 11. Example of method.—To illustrate the methods that we are now able to use, let us verify (V. 10. 4) by means of (V. 10. 1). It is given that

$$Y_\sigma = d_{\lambda\sigma} X^\lambda.$$

To find the value of $d^{\lambda\sigma} Y_\sigma$, it will not do to replace Y_σ by the above value as it stands, since we should then have

three λ's. We must first replace λ in the expression for Y_σ by some other suffix, say μ. We then have, by (V. 10. 1) and (V. 9. 5),

$$d^{\lambda\sigma} Y_\sigma = d^{\lambda\sigma} d_{\mu\sigma} X^\mu = 1^\lambda_\mu X^\mu = X^\lambda,$$

which is what we wanted to prove. The reader will find it instructive, for comparison, to write out the proof in the ordinary notation.

In the above proof we have proceeded from $d^{\lambda\sigma}(d_{\mu\sigma} X^\mu)$ to $(d^{\lambda\sigma} d_{\mu\sigma}) X^\mu$. It has already been pointed out, in § 6 (v), that summations in a case of this kind can be made in any order.

SETS

VI. SETS OF QUANTITIES

VI. 1. Introductory.—(i) It may have been noticed that in V. 11, in deducing (V. 10. 4) from (V. 10. 1) and (V. 9. 5), we made no direct use of determinant properties: the only indirect use being in the relations between elements and their cofactors, from which (V. 10. 1) was derived. But we can dispense even with this indirect use. In the equations $d_{\lambda\sigma} X^{\lambda} = Y_{\sigma}$ the values of $d_{\lambda\sigma}$ are supposed to be known; and we can treat the statement in (V. 10. 1), namely

$$d^{\lambda\sigma} d_{\mu\sigma} = |^{\lambda}_{\mu},$$

as a set of equations giving the values of $d^{\lambda\sigma}$ in terms of those of $d_{\lambda\sigma}$. If, for instance, $m = 20$, the set $d_{\lambda\sigma}$ contains 400 elements, and (V. 10. 1) is a condensed statement of the 400 equations (each with 20 terms on one side) which give the 400 values of $d^{\lambda\sigma}$. Thus for $p = 2$ we should have

$$\left.\begin{aligned} d_{11}d^{21} + d_{12}d^{22} + d_{13}d^{23} + \ldots + d_{1m}d^{2m} &= 0 \\ d_{21}d^{21} + d_{22}d^{22} + d_{23}d^{23} + \ldots + d_{2m}d^{2m} &= 1 \\ d_{31}d^{21} + d_{32}d^{22} + d_{33}d^{23} + \ldots + d_{3m}d^{2m} &= 0 \\ &\;\vdots \\ d_{m1}d^{21} + d_{m2}d^{22} + d_{m3}d^{23} + \ldots + d_{mm}d^{2m} &= 0 \end{aligned}\right\},$$

which give the values of $d^{21}, d^{22}, d^{23} \ldots d^{2m}$, i.e. of $d^{2\sigma}$. Similarly for $d^{1\sigma}, d^{3\sigma}$, etc.

(ii) We have, in fact, arrived at a position similar to that reached at the end of the first chapter. We started with the problem of solving a set of simultaneous equations, and arrived at a probable solution, involving what we

called determinants. To verify the solution, we had to investigate the properties of determinants. The determinant thus took the leading place, its applicability to the solving of equations being one only of its properties. A determinant of order m is based on a set of m^2 quantities, which for convenience of reference are thought of as arranged in a square, the determinant being expressed by enclosing the set of symbols of the quantities between vertical lines: and we have reached the stage at which the set of quantities becomes the important thing, its existence as the basis of a determinant being one only of its properties.

(iii) These properties we have now to consider. The following sections of this chapter are mainly a restatement, with obvious modifications and extensions, of results obtained in the preceding chapter.

VI. 2. Single sets.—(i) We may have a **single set**

$$A_\rho \equiv (A_1 \ \ A_2 \ \ A_3 \dots A_m).$$

The separate quantities $A_1 A_2 \dots A_m$ comprised in the set are called its **elements**. The **order** of the set is the number of elements comprised in it. The typical statement with regard to such a set is of the form

$$A_\rho = E_\rho.$$

This, in the first instance, we regarded merely as a short way of saying that

$$A_1 = E_1, \ A_2 = E_2, \dots A_m = E_m;$$

but we must now think of it as a statement that the two sets A_ρ and E_ρ, each taken as a whole, are equal, this equality of the wholes implying the equality of corresponding elements. The analogy of a vector may help us here.

The statement that two vectors are equal implies that the components are equal, each to each; but what we really think of is not the separate equalities of the components but the equality, in all respects, of the vectors.

(ii) Single sets behave like ordinary vectors as regards addition and subtraction and multiplication by a scalar (§ 4 (iii)), but not as regards multiplication of one single set by another.

VI. 3. Double sets.—(i) We may have a **double set** of order m—i.e. comprising m^2 elements—

$$A_{\mu\rho} \equiv \begin{pmatrix} A_{11} & A_{21} & A_{31} \ldots A_{m1} \\ A_{12} & A_{22} & A_{32} \ldots A_{m2} \\ \vdots & \vdots & \vdots \quad\;\; \vdots \\ A_{1m} & A_{2m} & A_{3m} \ldots A_{mm} \end{pmatrix}$$

Here the quantity in the qth column and rth row is A_{qr}. We adopt the convention that the first Greek letter (in alphabetical order) represents the column and the second the row, so that

$$A_{\rho\mu} \equiv \begin{pmatrix} A_{11} & A_{12} & A_{13} \ldots A_{1m} \\ A_{21} & A_{22} & A_{23} \ldots A_{2m} \\ \vdots & \vdots & \vdots \quad\;\; \vdots \\ A_{m1} & A_{m2} & A_{m3} \ldots A_{mm} \end{pmatrix},$$

the representative element of which is A_{rq}. The sets $A_{\mu\rho}$ and $A_{\rho\mu}$ are called the **transposed** of each other.

(ii) The determinant

$$| A_{qr} |$$

is called the determinant of the set* $A_{\mu\rho}$, and similarly $| A_{rq} |$ is the determinant of $A_{\rho\mu}$.

* The set is usually called the *matrix* of the determinant. It is a singularly inappropriate name, as the symbol of the set is the inner part of that of the determinant, not something which surrounds it. The set is really the substance or *core* of the determinant.

(iii) The brackets in which the elements of A_ρ, $A_{\mu\rho}$, $A_{\rho\mu}$ have been placed are not essential,* and have been introduced partly to help the eye and partly to indicate that the sets are being considered as a whole.

(iv) A double set is **symmetrical** if columns and rows can be interchanged without altering it. Thus, if $A_{\mu\rho}$ is symmetrical, then $A_{\mu\rho}=A_{\rho\mu}$: and conversely.

VI. 4. Sets generally.

—(i) We describe a single set as being of **rank**† 1, and a double set as being of rank 2.

(ii) Similarly a set of rank 3 of order m is made up of m double sets of order m; and so on. Thus we might represent a set of rank 3 by $A_{\mu\nu\rho}$. There would have to be a convention as to the order of the symbols, so as to distinguish $A_{\mu\nu\rho}$ from $A_{\nu\mu\rho}$ etc. Where, however, the set is symmetrical, so that $A_{\mu\nu\rho}=A_{\nu\mu\rho}=$etc., this difficulty does not arise.

(iii) The set of rank 0 is a single quantity or **scalar.**

(iv) To denote a set generally, without reference to its suffixes, we use a Gothic letter such as $\mathfrak{A}$ or $\mathfrak{B}$.

* An alternative method, in the case of a double set or matrix, is to enclose the symbols between two pairs of vertical lines, so as to distinguish the set from the determinant, which has two single lines. It is not a satisfactory symbolism from our point of view, as it would seem to suggest that the set is more restricted than the determinant, whereas what we are aiming at is to free the set from the bonds of the determinant.

† I have been doubtful as to the appropriate word. In reference to tensors Einstein uses *Rang*, Hilbert *Ordnung*, Eddington *rank*, de Sitter *order*. The objection to either of the last two is that there is already a settled meaning for *order* as regards a determinant, and (though this is not so important) for *rank* as regards a matrix (see note to § 3 (ii)). It would seem reasonable to describe a set containing m^f elements (i. e. composed of m sets each containing m^{f-1} elements) as being of *degree f*. I have, however, felt bound to keep to Eddington's use of *rank*.

VI. 5. **Sums and products of sets.**—(i) If two or more sets, of the same rank and the same order, have *the same* suffixes, we add (or subtract) them by adding (or subtracting) corresponding elements. Thus

$$A_\lambda + B_\lambda \equiv (A_1 + B_1 \quad A_2 + B_2 \quad A_3 + B_3 \ldots A_m + B_m);$$

and similarly for sets of higher rank.

(ii) We multiply a set, of whatever rank, by a scalar when we multiply every element of the set by the scalar; e.g.

$$kA_{\mu\rho} \equiv \begin{pmatrix} kA_{11} & kA_{21} & kA_{31} \ldots kA_{m1} \\ kA_{12} & kA_{22} & kA_{32} \ldots kA_{m2} \\ \vdots & \vdots & \vdots \quad\quad \vdots \\ kA_{1m} & kA_{2m} & kA_{3m} \ldots kA_{mm} \end{pmatrix}.$$

Thus the determinant of $kA_{\mu\rho}$ is not k times, but k^m times, the determinant of $A_{\mu\rho}$.

(iii) If $\mathfrak{A}$ and $\mathfrak{B}$ are two sets, with *different* suffixes, of ranks f and g respectively (f and g not being necessarily different), their **product** $\mathfrak{A}\,\mathfrak{B}$ is the set of rank $f+g$ obtained by multiplying every element of one by every element of the other. (Here, as elsewhere, we assume that all the sets we are considering are of the same order.) Thus the product of two single sets A_ρ and B_σ is the double set $A_\rho B_\sigma$ obtained by giving to ρ and σ separately each of the values **1** to m. A product obtained in this way is sometimes called an **outer** product, to distinguish it from an 'inner' product as defined in § 6 below.

VI. 6. **Inner product.**—(i) When a suffix occurs twice in an expression such as $A_{\nu\nu}$ or $B_\nu C_\nu$, or, more generally, in any single expression or product, e.g. $A_{\lambda\nu\nu\rho}\ldots$ or $B_{\nu\rho}\ldots C_{\nu\sigma}\ldots$ (where the letters may be in any order), this means that the expression is to be summed for the values 1, 2...m of the suffix; e.g.

$$\left.\begin{aligned} B_\nu C_\nu &\equiv B_1C_1 + B_2C_2 + \ldots + B_mC_m \\ B_\nu D_{\lambda\nu} &\equiv B_1D_{\lambda1} + B_2D_{\lambda2} + \ldots + B_mD_{\lambda m} \\ &\text{etc.} \end{aligned}\right\} . \qquad \textbf{(VI. 6. A)}$$

Where a letter occurs twice in this way, each of the two letters is **linked.** Where a letter occurs once only, it is **free.** The linked suffixes are called **dummy**, as they can be replaced by a pair of any other suffixes not already occurring in the expression.

(ii) In the particular case where the expression to be summed is of the form $B_\nu C_\nu$, the result is called the **inner product** of B_ν and C_ν, or of B_λ and C_ρ, etc. It is immaterial what suffixes we use in this latter description, since they have to be replaced by one and the same suffix.

(iii) From a pair of double sets $A_{\mu\rho}$ and $B_{\mu\rho}$, or $A_{\lambda\sigma}$ and $B_{\mu\rho}$, we can by a single product-summation form several different double sets $A_{\mu\nu}\, B_{\nu\rho}$, $A_{\lambda\rho}\, B_{\mu\rho}$, $A_{\lambda\mu}\, B_{\lambda\rho}$, etc. There is also the scalar quantity $A_{\mu\rho}\, B_{\mu\rho}$, formed by two summations, which can be simultaneous or successive; if successive, the first is a product-summation giving us the double set $A_{\lambda\rho}\, B_{\mu\rho}$ or $A_{\mu\rho}\, B_{\mu\sigma}$. Strictly speaking, this scalar quantity $A_{\mu\rho}\, B_{\mu\rho}$ is *the* inner product of $A_{\mu\rho}$ and $B_{\mu\rho}$. But it occurs less frequently than the double sets obtained by a single summation, and it is therefore more convenient to call one of these latter the inner product. We shall call $A_{\mu\rho}\, B_{\mu\rho}$ the **complete inner product** of $A_{\mu\rho}$ and $B_{\mu\rho}$. By analogy with the expression found in V. 6 (vi) for the product of two determinants, we define the **inner product*** of two sets $A_{\mu\rho}$ and $B_{\mu\rho}$—or, more generally, of

* This is what, in the case of matrices, is called the 'product'. The true product of two sets $A_{\mu\sigma}$ and $B_{\lambda\rho}$ is a set of rank 4.

If this use of 'inner product' seemed likely to lead to confusion with the 'complete inner product', we could use a different phrase, such as 'interproduct'. It should be noted that the inter-

two sets $A_{\mu\sigma}$ and $B_{\lambda\rho}$—, in each of which the letters are in their proper alphabetical order, as the set $A_{\mu\nu}B_{\nu\rho}$; the linked suffix in this latter expression being chosen so as to be (alphabetically) intermediate between the two free suffixes. This is equivalent to saying, as regards this case, that *the inner product of two double sets is the double set whose element in the* q*th column and* r*th row is the inner product of the* q*th column of the first set and the* r*th row of the second set*; and we apply this rule to all cases. The simplest way of applying it is to use the result for $A_{\mu\rho}$ and $B_{\mu\rho}$ and alter the order of the suffixes where necessary. Suppose, for instance, that we want the inner product of $A_{\mu\rho}$ and $B_{\rho\mu}$; then by writing $B_{\rho\mu} \equiv F_{\mu\rho}$ we see that the inner product is $A_{\mu\nu}F_{\nu\rho} = A_{\mu\nu}B_{\rho\nu}$. It should be noticed that in all cases the inner product depends on the relative position of the original sets; thus the inner product of $B_{\mu\rho}$ and $A_{\mu\rho}$ is not $A_{\mu\nu}B_{\nu\rho}$ but $B_{\mu\nu}A_{\nu\rho}$.

(iv) There are four main forms of inner product constructed in accordance with (iii); and four others, which are really repetitions, can be obtained by interchanging the two sets. Denoting the inner product of $\mathfrak{A}$ and $\mathfrak{B}$ (in this order) by $\mathfrak{A} \times \mathfrak{B}$ (cf. IV. 5 (ii) as to product of two determinants), the forms are as follows:

$$
\begin{array}{llll}
A_{\mu\rho} \times B_{\mu\rho} = A_{\mu\nu}B_{\nu\rho} & (1) & B_{\mu\rho} \times A_{\mu\rho} = B_{\mu\nu}A_{\nu\rho} = A_{\nu\rho}B_{\mu\nu} & (5) \\
A_{\mu\rho} \times B_{\rho\mu} = A_{\mu\nu}B_{\rho\nu} & (2) & B_{\rho\mu} \times A_{\mu\rho} = B_{\nu\mu}A_{\nu\rho} = A_{\nu\rho}B_{\nu\mu} & (6) \\
A_{\rho\mu} \times B_{\mu\rho} = A_{\nu\mu}B_{\nu\rho} & (3) & B_{\mu\rho} \times A_{\rho\mu} = B_{\mu\nu}A_{\rho\nu} = A_{\rho\nu}B_{\mu\nu} & (7) \\
A_{\rho\mu} \times B_{\rho\mu} = A_{\nu\mu}B_{\rho\nu} & (4) & B_{\rho\mu} \times A_{\rho\mu} = B_{\nu\mu}A_{\rho\nu} = A_{\rho\nu}B_{\nu\mu} & (8)
\end{array}
$$

(v) The transposed of an inner product such as $A_{\mu\nu}B_{\nu\rho}$ is found in the usual way (V. 5 (iii)) by interchanging the free suffixes μ and ρ. By comparison of (1) with (8), (2) with (7), etc., it will be seen that *the transposed of the inner product of two double sets is the*

mediate product-sum, for $A_{\mu\rho}B_{\mu\rho}$, is not $A_{\mu\nu}B_{\nu\rho}$ but either $A_{\lambda\rho}B_{\mu\rho}$ or $A_{\mu\rho}B_{\mu\sigma}$. There are various reasons for taking the former, rather than one of the two latter, as 'the' inner product.

inner product of the transposed sets, in reverse order; e.g. the transposed of the inner product of $A_{\rho\mu}$ and $B_{\rho\mu}$ is the inner product of $B_{\mu\rho}$ and $A_{\mu\rho}$.

(vi) It will be seen by comparison with IV. 5 (ii) and the Appendix that the rule for construction of the inner product of two double sets is exactly the same as that for construction of the product of two determinants; so that the determinant of the inner product of two double sets—whether we call them (say) $A_{\mu\rho}$ and $B_{\rho\mu}$ or $A_{\mu\sigma}$ and $B_{\rho\lambda}$—is equal to the product of the determinants of the two sets.

VI. 7. The unit set.

—(i) The **unit set***

$$|_{\rho}^{\mu}$$

is defined as the set whose typical term is

$$|_{r}^{q} \equiv \begin{Bmatrix} 1 \text{ if } r = q \\ 0 \text{ if } r \neq q \end{Bmatrix}, \qquad \textbf{(VI. 7. A)}$$

so that

$$|_{\rho}^{\mu} \equiv \begin{pmatrix} 1 & 0 & 0 & 0\ldots0 \\ 0 & 1 & 0 & 0\ldots0 \\ 0 & 0 & 1 & 0\ldots0 \\ \vdots & \vdots & \vdots & \vdots \quad \vdots \\ 0 & 0 & 0 & 0\ldots1 \end{pmatrix}. \qquad \textbf{(VI. 7. 1)}$$

(ii) From the definition it follows that the set is symmetrical, i.e.

$$|_{\rho}^{\mu} = |_{\mu}^{\rho}, \qquad \textbf{(VI. 7. 2)}$$

and that

$$\big||_{r}^{q}\big| = 1. \qquad \textbf{(VI. 7. 3)}$$

(iii) The special property of this set is that, if A_{μ} is any set (possibly containing other suffixes $\sigma\ \tau\ldots$), then

$$|_{\mu}^{\nu} A_{\mu} = |_{\nu}^{\mu} A_{\mu} = A_{\nu}, \qquad \textbf{(VI. 7. 4)}$$

so that the inner product of the unit set and any other set

* This is by analogy with the 'unit matrix'.

is the same as the latter but with the suffix changed. In other words, the unit set acts, for inner multiplication, as a *substitution-operator*.

(iv) In particular, the inner product of two unit sets is a unit set, i.e.

$$|^{\mu}_{\nu} |^{\nu}_{\rho} = |^{\mu}_{\rho}. \qquad \text{(VI. 7. 5)}$$

VI. 8. Inverse double sets.—(i) We take any double set

$$A_{\mu\rho},$$

and we say that there is another set

$$A^{\rho\mu}$$

connected with it by the condition that the inner product of the former and the latter is a unit set, i.e. (see §§ 6 (iii) and 7 (i)) that

$$A_{\mu\nu} A^{\rho\nu} = |^{\rho}_{\mu}. \qquad \text{(VI. 8. 1)}$$

This represents m^2 equations, which are sufficient to determine the m^2 values of A^{rq} when those of A_{qr} are known. The set $A^{\rho\mu}$, as defined by the above condition, is called the **inverse** of the set $A_{\mu\rho}$. We shall keep to this notation, so that (VI. 8. 1) will always hold, however we alter the letters A, μ, ν, ρ.

(ii) The above is subject to one condition. If we write down the equations which determine the elements in, say, the second row of $A^{\rho\mu}$, namely

$$\left.\begin{aligned} A_{11}A^{21} + A_{12}A^{22} + A_{13}A^{23} + \dots + A_{1m}A^{2m} &= 0 \\ A_{21}A^{21} + A_{22}A^{22} + A_{23}A^{23} + \dots + A_{2m}A^{2m} &= 1 \\ A_{31}A^{21} + A_{32}A^{22} + A_{33}A^{23} + \dots + A_{3m}A^{2m} &= 0 \\ &\vdots \\ A_{m1}A^{21} + A_{m2}A^{22} + A_{m3}A^{23} + \dots + A_{mm}A^{2m} &= 0 \end{aligned}\right\},$$

we see that in order that there may be a solution it is necessary that we should have $|A_{rq}| \neq 0$, which is the same thing as

$$|A_{qr}| \neq 0. \qquad \text{(VI. 8. 2)}$$

This applies also to the other rows. It is a sufficient as well as a necessary condition for the existence of $A^{\rho\mu}$.

(iii) Taking it that $|A_{qr}| \neq 0$, we have, by (V. 6. 2) and (VI. 8. 1) and (VI. 7. 3),

$$|A^{qr}| \times |A_{rq}| = |A^{q\nu} A_{r\nu}| = \| {}^q_r \| = 1. \qquad \text{(VI. 8. 3)}$$

It follows that

$$|A^{qr}| \neq 0, \quad |A^{rq}| \neq 0. \qquad \text{(VI. 8. 4)}$$

(iv) The statement (VI. 8. 1) is a statement as to the m^2 relations obtained by taking each value of ρ with each value of μ. It is therefore equally true to say, by interchanging μ and ρ, that

$$A^{\mu\nu} A_{\rho\nu} = |{}^\mu_\rho. \qquad \text{(VI. 8. 5)}$$

The expression on the left-hand side is (§ 6 (iii)) the inner product of $A^{\mu\rho}$ and $A_{\rho\mu}$; and these are the transposed of $A^{\rho\mu}$ and $A_{\mu\rho}$ respectively. Hence, if $\mathfrak{B}$ is the inverse of $\mathfrak{A}$, the transposed of $\mathfrak{A}$ is the inverse of the transposed of $\mathfrak{B}$.

(v) The relation in (VI. 8. 1) is a relation connecting columns of the original set and rows of the inverse set. There is a similar relation connecting rows and columns. For (VI. 8. 1) gives

$$A^{\mu\lambda} A_{\mu\nu} A^{\rho\nu} = |{}^\rho_\mu A^{\mu\lambda} = A^{\rho\lambda} = |{}^\lambda_\nu A^{\rho\nu};$$

whence, as will be shown in § 9 (v), it follows that

$$A^{\mu\lambda} A_{\mu\nu} = |{}^\lambda_\nu, \qquad \text{(VI. 8. 6)}$$

and hence also

$$A_{\mu\lambda} A^{\mu\nu} = |{}^\nu_\lambda. \qquad \text{(VI. 8. 7)}$$

The expression on the left-hand side of (VI. **8. 6**) is the inner product of $A^{\nu\lambda}$ and $A_{\lambda\nu}$, which is the equivalent of that of $A^{\rho\mu}$ and $A_{\mu\rho}$; so that, by the definition in (i), the latter is the inverse of the former. Hence, if $\mathfrak{B}$ is the inverse of $\mathfrak{A}$, then $\mathfrak{A}$ is the inverse of $\mathfrak{B}$. The relation (VI. **8. 7**) is similar to (VI. **8. 5**), and shows that the transposed of $\mathfrak{B}$ is the inverse of the transposed of $\mathfrak{A}$.

VI. 9. Reciprocation.—(i) Suppose there are two single sets X^λ and Y_λ connected by the relation

$$Y_\sigma = A_{\lambda\sigma} X^\lambda.$$

Then, as in V. **11**, we have

$$A^{\lambda\sigma} Y_\sigma = A^{\lambda\sigma} A_{\mu\sigma} X^\mu = \mathsf{I}^{\lambda}_{\mu} X^\mu = X^\lambda;$$

i. e.

$$\text{If} \quad Y_\sigma = A_{\lambda\sigma} X^\lambda, \ \text{then} \ X^\lambda = A^{\lambda\sigma} Y_\sigma. \qquad \textbf{(VI. 9. 1)}$$

Similarly by taking X^λ to be each single set, in turn, of a set of second or higher rank, with Y_λ to correspond, we find that

$$\text{If} \ Y_{\sigma\tau\ldots} = A_{\lambda\sigma} X^\lambda_{\tau\ldots}, \ \text{then} \ X^\lambda_{\tau\ldots} = A^{\lambda\sigma} Y_{\sigma\tau\ldots}. \qquad \textbf{(VI. 9. 2)}$$

(ii) Thus the operation represented by $A_{\lambda\sigma}$ is annulled by the operation $A^{\lambda\sigma}$; and conversely. The sets $A_{\lambda\sigma}$ and $A^{\lambda\sigma}$ will be said to be **reciprocal** to one another: and the process adopted in (VI. 9. **1**) and (VI. 9. **2**)—which we shall have to use very frequently—will be called **reciprocation.**

(iii) We see from § 8 that the reciprocal of a set is the transposed of the inverse of the set, and conversely. If a set is symmetrical, its inverse and its reciprocal are identical.

(iv) As an example of the application of (VI. 9. 1), suppose that

$$A_{\lambda\sigma} X^\lambda = 0$$

for all values of σ. Then, by reciprocation,

$$X^{\lambda} = A^{\lambda\sigma}\, 0 = 0,$$

provided that $|A_{qr}| \neq 0$. (This is practically the same thing as saying that, if the inner products of X^{λ} by m independent single sets are all 0, then X^{λ} is 0.)

(v) Similarly, suppose that

$$A_{\lambda\nu}\, C^{\nu\sigma} = A_{\lambda\nu}\, D^{\nu\sigma}$$

identically, i.e. for all values of λ and σ, and that $|A_{qr}|$ is not $= 0$. Then, by reciprocation,

$$C^{\nu\sigma} = A^{\lambda\nu} A_{\lambda\rho}\, D^{\rho\sigma} = |^{\nu}_{\rho}\, D^{\rho\sigma} = D^{\nu\sigma}.$$

Thus we can divide both sides of the equation by $A_{\lambda\nu}$. This supplies the missing step in § 8 (v).

(vi) The two definitions, and the proposition, used in the establishment of (VI. 9. 1) are

$$|^{q}_{r} = \begin{Bmatrix} 1 \text{ if } r = q \\ 0 \text{ if } r \neq q \end{Bmatrix}, \quad . \quad . \quad . \quad . \quad . \quad \text{(A)}$$

$$|^{\lambda}_{\mu}\, A^{\mu} = A^{\lambda}, \quad . \quad . \quad . \quad . \quad . \quad \text{(1)}$$

$$A_{\mu\sigma} A^{\lambda\sigma} = |^{\lambda}_{\mu}; \quad . \quad . \quad . \quad . \quad . \quad \text{(B)}$$

and from these we deduce (VI. 9. 1). We could have altered the order in various ways. For instance, we could have defined $|^{\lambda}_{\mu}$ by (1); thence, by giving λ its successive values, and equating coefficients, we should have got (A). Also we might have defined $A^{\lambda\sigma}$ by (VI. 9. 1), instead of by (B), as the coefficients of Y_{σ} when the equations $Y_{\sigma} = A_{\lambda\sigma} X^{\lambda}$ are solved for X^{λ}. This would give $X^{\lambda} = A^{\lambda\sigma} A_{\mu\sigma} X^{\mu}$. Then, if we defined $|^{\lambda}_{\mu}$ by (1), we should have $A^{\lambda\sigma} A_{\mu\sigma} = |^{\lambda}_{\mu}$; or, if we defined $|^{\lambda}_{\mu}$ by (B), we should have $X^{\lambda} = |^{\lambda}_{\mu}\, X^{\mu}$, which is (1).

VI. 10. **Continued inner products.**—(i) We can construct continued inner products without ambiguity, provided we adhere to the rule laid down in § 6 (iii). Suppose, for instance, that we

want the inner product of $A_{\lambda\sigma}$, $B_{\lambda\sigma}$, and $C_{\sigma\lambda}$. That of $A_{\lambda\sigma}$ and $B_{\lambda\sigma}$, according to the rule, is $A_{\lambda\mu}B_{\mu\sigma}$. If we call this $F_{\lambda\sigma}$, then, by (2) of § 6 (iv), the inner product of $F_{\lambda\sigma}$ and $C_{\sigma\lambda}$ is $F_{\lambda\nu}C_{\sigma\nu}$; i.e. the inner product of $A_{\lambda\sigma}$, $B_{\lambda\sigma}$, and $C_{\sigma\lambda}$ is $A_{\lambda\mu}B_{\mu\nu}C_{\sigma\nu}$. Similarly that of $A_{\lambda\sigma}$, $B_{\lambda\sigma}$, $C_{\sigma\lambda}$, and $D_{\lambda\sigma}$ is $A_{\lambda\mu}B_{\mu\nu}C_{\rho\nu}D_{\rho\sigma}$.

(ii) On the other hand, the inner product of $A_{\sigma\lambda}$, $B_{\sigma\lambda}$, $C_{\lambda\sigma}$, $D_{\sigma\lambda}$ is not $A_{\sigma\rho}B_{\rho\nu}C_{\mu\nu}D_{\mu\lambda}$. For, by (4) of § 6 (iv), that of $A_{\sigma\lambda}$ and $B_{\sigma\lambda}$ is $B_{\sigma\mu}A_{\mu\lambda}$. Calling this $G_{\sigma\lambda}$, the inner product of $G_{\sigma\lambda}$ and $C_{\lambda\sigma}$ is, by (3) of § 6 (iv), $C_{\nu\sigma}G_{\nu\lambda} = C_{\nu\sigma}B_{\nu\mu}A_{\mu\lambda}$. Similarly that of $A_{\sigma\lambda}$, $B_{\sigma\lambda}$, $C_{\lambda\sigma}$, $D_{\sigma\lambda}$ is $D_{\sigma\rho}C_{\nu\rho}B_{\nu\mu}A_{\mu\lambda}$.

(iii) The transposed of the inner product of any number of double sets is the inner product of the transposed sets, in reverse order; e.g. the transposed of the inner product of $A_{\lambda\sigma}$, $B_{\lambda\sigma}$, $C_{\sigma\lambda}$, $D_{\lambda\sigma}$ is the inner product of $D_{\sigma\lambda}$, $C_{\lambda\sigma}$, $B_{\sigma\lambda}$, $A_{\sigma\lambda}$. [For we have shown, in § 6 (v), that this is true for the inner product of two sets; and thence it follows, by induction, for any number of sets.]

(iv) The inverse of the inner product of any number of double sets is the inner product of the inverse sets, in reverse order; e.g. the inverse of the inner product of $B_{a\sigma}$, $C_{a\sigma}$, $D_{a\sigma}$, $E_{a\sigma}$ is the inner product of $E^{\sigma a}$, $D^{\sigma a}$, $C^{\sigma a}$, $B^{\sigma a}$, i.e.

If $A_{a\sigma} \equiv B_{a\beta}C_{\beta\gamma}D_{\gamma\delta}E_{\delta\sigma}$, then $A^{\sigma a} = B^{\sigma\delta}C^{\delta\gamma}D^{\gamma\beta}E^{\beta a}$. **(VI.10.1)**

[Denote this latter expression (right-hand side) by $F^{\sigma a}$, and alter $\beta\ \gamma\ \delta$ in it to $\mu\ \nu\ \rho$. Then the inner product of $A_{a\sigma}$ and $F^{\sigma a}$ is

$$\begin{aligned} A_{a\lambda}F^{\sigma\lambda} &= B_{a\beta}B^{\sigma\rho}C_{\beta\gamma}C^{\rho\nu}D_{\gamma\delta}D^{\nu\mu}E_{\delta\lambda}E^{\mu\lambda} \\ &= B_{a\beta}B^{\sigma\rho}C_{\beta\gamma}C^{\rho\nu}D_{\gamma\delta}D^{\nu\mu}\big|^{\mu}_{\delta} \\ &= B_{a\beta}B^{\sigma\rho}C_{\beta\gamma}C^{\rho\nu}\big|^{\nu}_{\gamma} \\ &= B_{a\beta}B^{\sigma\rho}\big|^{\rho}_{\beta} \\ &= \big|^{\sigma}_{a}. \end{aligned}$$

Hence, by reciprocation,

$$F^{\sigma\lambda} = A^{a\lambda}\big|^{\sigma}_{a} = A^{\sigma\lambda},$$

so that

$$F^{\sigma a} = A^{\sigma a}.]$$

(v) It follows from (iv) and (iii), since $A^{a\sigma}$ is the transposed of $A^{\sigma a}$, that the reciprocal of the inner product of any number of double sets is the inner product of the reciprocal sets; e.g.

If $A_{a\sigma} \equiv B_{a\beta}C_{\beta\gamma}D_{\gamma\delta}E_{\delta\sigma}$, then $A^{a\sigma} = B^{a\beta}C^{\beta\gamma}D^{\gamma\delta}E^{\delta\sigma}$.

(VI. 10. 2)

VI. 11. **Partial sets.**—(i) When we are dealing with a set

$$A_\lambda \equiv (A_1 \ A_2 \ A_3 \ldots A_m)$$

we sometimes want to consider the separate or mutual relations of groups of the A's. The simplest case is when the set divides into two groups, one group consisting of k elements, which we shall take to be the first k, and the other group consisting of the other $m-k$ elements. If we use suffixes $\alpha\,\beta\,\gamma\ldots$ in reference to the first group, and $\phi\,\chi\,\psi\ldots$ in reference to the second, reserving $\lambda\,\mu\,\nu\ldots$ for the set as a whole, we may treat the two groups as **partial single sets** of orders k and $m-k$ respectively, and write

$$A_\alpha \equiv (A_1 \ A_2 \ldots A_k), \ A_\phi \equiv (A_{k+1} \ A_{k+2} \ldots A_m).$$

(ii) In the same way a double set $A_{\mu\rho}$ may fall into four groups by division by two lines cutting off k columns and k rows respectively. We could denote these groups by

$$\begin{Bmatrix} A_{\alpha\gamma} & A_{\phi\gamma} \\ A_{\alpha\psi} & A_{\phi\psi} \end{Bmatrix},$$

ϕ being regarded as coming before γ. The groups $A_{\alpha\gamma}$ and $A_{\phi\psi}$ would be **partial double sets** of orders k and $m-k$ respectively. The groups $A_{\alpha\psi}$ and $A_{\phi\gamma}$ would each have different numbers of columns and of rows, and therefore would not be double sets; but this would usually not matter, as we should be specially concerned with $A_{\alpha\gamma}$ and $A_{\phi\psi}$. The important point to notice is that, if we take $A_{\alpha\gamma}$, say, as a partial set and construct the inverse set $A^{\gamma\alpha}$ or the reciprocal set $A^{\alpha\gamma}$, the set so constructed will not in general be the same as the set made up of the corresponding elements of the set inverse or reciprocal to $A_{\mu\rho}$. The inverse set $A^{\gamma\alpha}$, for instance, is given by

$$A_{\alpha\beta}A^{\gamma\beta} = |^{\gamma}_{\alpha},$$

with summations made only from 1 to k instead of from 1 to m. To avoid mistake, we may write it $(A^{\gamma\alpha})_k$. Similarly the set inverse to the partial set $A_{\phi\psi}$ may be written $[A^{\psi\phi}]_{m-k}$.

(iii) If, however, all the elements in the portions $A_{\alpha\psi}$ and $A_{\phi\gamma}$ are 0, so that the set $A_{\mu\rho}$ practically consists only of the two partial sets $A_{\alpha\gamma}$ and $A_{\phi\psi}$, this is also the case for the complete reciprocal set $A^{\mu\rho}$; all the elements in $A^{\alpha\psi}$ and $A^{\phi\gamma}$ are 0, and the elements of $A^{\alpha\gamma}$ and $A^{\phi\psi}$ are just the same whether they are

regarded as obtained from the complete set $A_{\mu\rho}$ or from the partial sets $A_{a\gamma}$ and $A_{\phi\psi}$.

(iv) If we had two or more sets, single or double, divided in the manner described above, we could take portions from different sets to form new sets. If, for instance, we had divided A_λ into A_a and A_ϕ, and B_λ into B_a and B_ϕ (orders again k and $m-k$), we could construct a new set consisting of A_a and B_ϕ.

(v) It would, of course, be incorrect to describe this new set as being A_a+B_ϕ, or the old set as being A_a+A_ϕ: for we cannot add together two sets of different order. We could, however, look at the matter in another way. Consider the two sets

$$\begin{array}{l}(A_1 \quad A_2 \ldots A_k \quad 0 \quad 0 \quad \ldots 0\ ), \\ (0 \quad 0 \ldots 0 \quad A_{k+1} \quad A_{k+2} \ldots A_m).\end{array}$$

The sum of these, if we regard each as having the suffix λ, is A_λ; and in this sense, if we denote the two sets by A_a and A_ϕ, and regard a and ϕ as connoting λ, we could say that

$$A_\lambda = A_a + A_\phi.$$

If we compare A_λ with a vector, we see that A_a and A_ϕ correspond to the projections of A_λ on a 'plane', i.e. a surface of the first degree, passing through the first k axes, and on a 'plane' through the last $m-k$ axes, respectively.

The extreme form, if we made further divisions, would be that in which A_λ was split up into m component single sets, each having $m-1$ 0's in it. It would only be in this sense that we could describe the set as being the sum of its m components.

VII. RELATED SETS OF VARIABLES

VII. 1. **Variable sets.**—(i) In the earlier chapters we considered the manner in which determinants arose in solving a set of equations of the form

$$(s = 1, 2, \dots m)\, d_{1s}X^1 + d_{2s}X^2 + \dots + d_{ms}X^m = Y_s; \quad (1)$$

and in the chapter preceding this we have considered the general aspects of the system under which we express these equations and their solution in the form

$$Y_\sigma = d_{\lambda\sigma}X^\lambda, \quad X^\lambda = d^{\lambda\sigma}Y_\sigma. \quad (2)$$

According to the definitions we gave to the notation, X^λ and Y_σ are each used in different senses in the two places where they occur in (2): X^λ means 'the elements of X^λ' on its first occurrence and 'each element of X^λ' on its second occurrence; and similarly for Y_σ, but in the reverse order. We have, however, by this time practically reached the stage of treating a set as a whole, so that we can now regard (2) as a pair of statements, one of which gives an expression for the set Y_σ in terms of the set X^λ, while the other expresses X^λ in terms of Y_σ. The form of either expression determines the nature of the relation between the two sets.

(ii) The special features of the particular case were that the X's were unknown quantities which we wanted to find, that the d's were coefficients, more or less accidental, and that the Y's were known quantities arising from the application of these coefficients to the X's; and, more important,

that this was merely an isolated set of equations, for which we had no further use when we had found the X's.

(iii) The relations with which we have to deal in this and subsequent chapters are of a different nature. We have a set $\mathfrak{A}$ and a set $\mathfrak{B}$, each consisting of a number (the same for both) of elements, which we will call the A's and the B's. In each set the elements need not be all of the same kind. Each of the A's is a variable; i.e. it either has, or can (as in the theory of statistics or of error) be regarded as having, a very large number of actual or possible values. These variables, the values of which are algebraically independent,* together constitute the variable set $\mathfrak{A}$. In the same way the B's are variables, and constitute another variable set $\mathfrak{B}$. But the two sets of variables are not independent of each other: they are connected by certain relations, by means of which the B's are known if the A's are known, and conversely. Thus the B's are functions of the A's, in the ordinary sense of the word, and the A's are functions of the B's. In this case we say that $\mathfrak{B}$ is a function of $\mathfrak{A}$, and $\mathfrak{A}$ a function of $\mathfrak{B}$. But we must not only say it, but think it; i.e. we must treat the functional relations of the A's and the B's rather as interpreting the nature of the functionality of $\mathfrak{A}$ and $\mathfrak{B}$ than as actually constituting this functionality.

(iv) In the particular case we have been considering, $\mathfrak{A}$ and $\mathfrak{B}$ were the single sets X^{λ} and Y_{λ}, and the relation between them was *linear*; i.e. the Y's were linear functions of the X's, and the X's were therefore linear functions of

* By algebraical independence of m quantities we mean that each may have any of its values, whatever the values of the other $m-1$ may be. This does not imply statistical independence, which is a different thing.

the Y's. In such a case we say that Y_λ is a **linear function** of X^λ, and it follows that X^λ is a linear function of Y_λ.

(v) In dealing with the theory of the subject, as distinct from its applications, we are concerned not with the actual values of elements of sets but with the relations between the sets. Thus in the case of the linear relation $Y_\sigma = d_{\lambda\sigma}X^\lambda$, where the variable single set Y_σ is expressed in terms of the variable single set X^λ and the fixed double set $d_{\lambda\sigma}$, the elements of $d_{\lambda\sigma}$ form a kind of framework

$$\left.\begin{array}{l} \wedge = d_{11} \wedge + d_{21} \wedge + d_{31} \wedge + \ldots + d_{m1} \wedge \\ \wedge = d_{12} \wedge + d_{22} \wedge + d_{32} \wedge + \ldots + d_{m2} \wedge \\ \vdots \\ \wedge = d_{1m} \wedge + d_{2m} \wedge + d_{3m} \wedge + \ldots + d_{mm} \wedge \end{array}\right\}$$

into which the values of the X's and the Y's can be fitted; and what we are really investigating are the properties and mutual relations of such frameworks. In the present chapter we shall consider certain simple relations between two such frameworks, namely relations between the linear relation of one pair of sets and the linear relation of another pair of sets.

VII. 2. Direct proportion of single sets.—(i) If a quantity Z is a linear function of m X's, which we will call $X^1, X^2, \ldots X^m$, it is of the form

$$Z = h_1 X^1 + h_2 X^2 + \ldots + h_m X^m = h_\lambda X^\lambda. \tag{1}$$

This is the simplest form of statement of a linear relation. Suppose, for instance, that Z is the 3rd difference of the X's, formed in the usual way, i. e.

$$Z = \Delta\Delta\Delta X^1.$$

This is equivalent to $Z = X^4 - 3X^3 + 3X^2 - X^1$, so that

$$h_1 = -1,\ h_2 = 3,\ h_3 = -3,\ h_4 = 1,\ h_5 = h_6 = \ldots = h_m = 0.$$

The h's having these values, the X's may alter, but (1) will always give the 3rd difference.

(ii) Now suppose that there is another set A^λ, and that C is the same function of the A's that Z is of the X's; e.g., as in the above example, that it is the 3rd difference. Then the h's are the same, so that

$$C = h_\lambda A^\lambda. \tag{2}$$

We can write (1) in the form

$$Z/X^\lambda = h_\lambda,$$

on the understanding that a suffix in a denominator is linked with a similar suffix on the other side and implies an inner multiplication. Similarly we can write (2) as

$$C/A^\lambda = h_\lambda.$$

Equating the two values of h_λ, we have

$$\frac{C}{A^\lambda} = \frac{Z}{X^\lambda} \tag{3}$$

as our way of stating that C is the same linear function of the A's that Z is of the X's.

(iii) Next suppose that a set Y_ρ* is a linear function of the set X^λ, so that each of the Y's is a linear function of the X's. Then we can take Z of (1) to be each of the Y's in turn: but the sets of h's will be different, so that the relation will be of the form

$$Y_\rho = d_{\mu\rho} X^\mu.$$

If there is also a set B_ρ, each element of which is the same

* The set might be called either Y^ρ or Y_ρ; we choose the latter as giving a convenient symbol $d_{\mu\rho}$ for the coefficient of X^μ (see V. 10 (i)).

linear function of the A's that the corresponding element of Y_ρ is of the X's, then

$$B_\rho = d_{\mu\rho} A^\mu,$$

the two sets of d's being the same. We therefore have*

$$\frac{B_\rho}{A^\mu} = \frac{Y_\rho}{X^\mu}$$

as a statement of the fact that the B's have the same linear relations to the A's that the Y's have to the X's. This would be the case, for instance, if the Y's were the successive differences of the X's, and the B's were those of the A's according to the same system.

In view of the variety of ways in which we are able to deal with sets according to algebraical laws, it is perhaps permissible to describe this as a case of **direct proportion**, and to say that B_ρ bears the same **ratio** to A^μ that Y_ρ bears to X^μ. This 'ratio', here denoted by $d_{\mu\rho}$, is really the operator that is required to convert A^μ into B_ρ or X^μ into Y_ρ.

(iv) If the linear relation of the B's to the A's is the same as that of the Y's to the X's, then that of the A's to the B's is the same as that of the X's to the Y's (*or*, if the ratio of B_ρ to A^μ is the same as that of Y_ρ to X^μ, then the ratio of A^μ to B_ρ is the same as that of X^μ to Y_ρ); i.e.

$$\text{If } \frac{B_\rho}{A^\mu} = \frac{Y_\rho}{X^\mu}, \text{ then } \frac{A^\mu}{B_\rho} = \frac{X^\mu}{Y_\rho}. \qquad \textbf{(VII. 2. 1)}$$

[Let $B_\rho = d_{\mu\rho} A^\mu$. Then $A^\mu = d^{\mu\rho} B_\rho$. Similarly $X^\mu = d^{\mu\rho} Y_\rho$. Therefore $A^\mu / B_\rho = X^\mu / Y_\rho$.]

* This expression B_ρ/A^μ must not be confused with B_ρ/A^μ as the double set whose typical element is B_r/A^q. The limitation in V. 5 (vi) excludes double sets of this kind from consideration.

(v) The above sets of relations can be expressed by the diagram in Fig. 1. The crosses may be taken as representing either coefficients of the X's in the values of the Y's and of the A's in the values of the B's; or—in virtue of (VII. 2. 1)—coefficients of the Y's in the values of the X's and of the B's in the values of the A's.

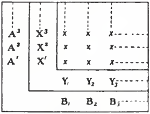


FIG. 1.

(vi) Ratios of the kind considered above can be combined according to the laws of ordinary algebra; e.g.

$$\frac{B_\nu}{A^\mu}\cdot\frac{C^\sigma}{B_\nu} = \frac{C^\sigma}{B_\nu}\cdot\frac{B_\nu}{A^\mu} = \frac{C^\sigma}{A^\mu}. \qquad \textbf{(VII. 2. 2)}$$

The expression on the left-hand side is, of course, an inner product. A special ratio is

$$\frac{A^\nu}{A^\mu} = |^{\nu}_{\mu} = |^{\mu}_{\nu} = \frac{A^\mu}{A^\nu}. \qquad \textbf{(VII. 2. 3)}$$

Example. If $B_\rho/A^\mu = Y_\rho/X^\mu$, prove that $B_\rho/A^\mu . X^\mu/Y_\sigma = |^{\sigma}_{\rho}$.

VII. 3. Reciprocal proportion of single sets.—(i) The other important class of cases is that in which the linear relation (or ratio) of B_ρ to A^μ is the *reciprocal* of that of Y_ρ to X^μ, i.e. in which, B_ρ and A^μ being altered to B^ρ and A_μ,

$$Y_\rho = k_{\mu\rho}X^\mu, \quad B^\rho = k^{\mu\rho}A_\mu.$$

This gives

$$X^\mu = k^{\mu\rho}Y_\rho,$$

so that

$$\frac{B^\rho}{A_\mu} = \frac{X^\mu}{Y_\rho}.$$

We can call this a case of **reciprocal proportion**.

(ii) If the ratio of B^ρ to A_μ is the reciprocal of that of Y_ρ to X^μ, then the ratio of Y_ρ to X^μ is the reciprocal of that of B^ρ to A_μ; i. e.

$$\text{If } \frac{B^\rho}{A_\mu} = \frac{X^\mu}{Y_\rho}, \text{ then } \frac{A_\mu}{B^\rho} = \frac{Y_\rho}{X^\mu}. \qquad \textbf{(VII. 3. 1)}$$

[Let $B^\rho = k^{\mu\rho} A_\mu$. Then $A_\mu = k_{\mu\rho} B^\rho$. Similarly $Y_\rho = k_{\mu\rho} X^\mu$. Therefore $A_\mu / B^\rho = Y_\rho / X^\mu$.]

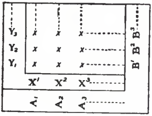


FIG. 2.

(iii) The sets of relations can be expressed by a diagram as in Fig. 2, where the crosses represent coefficients of the Y's in the X's and of the A's in the B's, or of the X's in the Y's and of the B's in the A's.

(iv) The inner products of the reciprocally corresponding sets are equal; i. e.

$$\text{If } \frac{B^\rho}{A_\mu} = \frac{X^\mu}{Y_\rho}, \text{ then } B^\rho Y_\rho = A_\mu X^\mu. \qquad \textbf{(VII. 3. 2)}$$

[Let $B^\rho = k^{\mu\rho} A_\mu$. Then $X^\mu = k^{\mu\rho} Y_\rho$; and therefore $Y_\rho = k_{\mu\rho} X^\mu$. Hence $B^\rho Y_\rho = k^{\mu\rho} A_\mu k_{\nu\rho} X^\nu = k^{\mu\rho} k_{\nu\rho} A_\mu X^\nu = |^\mu_\nu A_\mu X^\nu = A_\mu X^\mu$.]

(v) An interesting case is that in which the Y's are the successive differences of the X's. It will be found that in this case the B's are linear functions of successive sums, and therefore of successive moments, of the A's. In the ordinary system, for instance, which is such as to give $Y_1 = X^1$, $Y_2 = \Delta X^1 = X^2 - X^1$, $Y_3 = \Delta\Delta X^1 = X^3 - 2X^2 + X^1, \ldots$, we have $X^1 = Y_1$, $X^2 = Y_1 + Y_2$, $X^3 = Y_1 + 2Y_2 + Y_3, \ldots$; and these give $B^1 = -\Sigma A_1$, $B^2 = +\Sigma\Sigma A_2$, $B^3 = -\Sigma\Sigma\Sigma A_3, \ldots$, the constants in the sums being chosen so that $\Sigma A_{m+1} = 0$, $\Sigma\Sigma A_{m+1} = 0$, $\Sigma\Sigma\Sigma A_{m+1} = 0, \ldots$

VII. 4. Cogredience and contragredience.—(i) In §§ 2 and 3 we have not assumed the existence of any relation between A^λ and X^λ or between B_λ and Y_λ. Where a relation does exist, the important cases are those of *cogredience* and *contragredience*. We start with a set X^λ, and a set A^λ which is derived in a definite way from X^λ, in other words is a function of X^λ. We then take Y_λ to be any linear function of X^λ, and B_λ to be derived from Y_λ in the same way that A^λ is derived from X^λ. Then B_λ is some function of A^λ. Of the cases in which this is a linear function, we are concerned with two special classes :—

(1) If B_ρ/A^μ is always $= Y_\rho/X^\mu$, then A^λ and X^λ are said to be **cogredient**.

(2) If B_ρ/A^μ is always $= X^\mu/Y_\rho$, then A^λ and X^λ are said to be **contragredient**.

An example of cogredience is given in IX. 6 (x), and of contragredience in VIII. 3 (iv) and IX. 4 (v).

(ii) Instead of saying that A^λ and X^λ are cogredient or contragredient, we might say that A^λ in the one case *varies directly* as X^λ and in the other case *varies reciprocally* as X^λ. When we say that A^λ **varies directly** as X^λ, we mean that, if X^λ is multiplied by any double set involving λ, A^λ is multiplied by the same set: when we say that A^λ **varies reciprocally** as X^λ, we mean that, if X^λ is multiplied by any double set involving λ, A^λ is multiplied by the reciprocal of this set.

(iii) If A^λ and X^λ are $\left\{\begin{matrix}\text{cogredient}\\ \text{contragredient}\end{matrix}\right\}$, and P^λ and A^λ are also $\left\{\begin{matrix}\text{cogredient}\\ \text{contragredient}\end{matrix}\right\}$, then P^λ and X^λ are cogredient; if A^λ and X^λ are $\left\{\begin{matrix}\text{cogredient}\\ \text{contragredient}\end{matrix}\right\}$, but P^λ and A^λ are $\left\{\begin{matrix}\text{contragredient}\\ \text{cogredient}\end{matrix}\right\}$, then P^λ and X^λ are contragredient.

[Suppose, for instance, that A^λ and X^λ are contragredient, and P^λ and A^λ are also contragredient. Then we have relations of the form

$$B_\rho/A^\mu = X^\mu/Y_\rho, \quad Q_\rho/P^\mu = A^\mu/B_\rho,$$

whence, by (VII. 3. 1),

$$Q_\rho/P^\mu = Y_\rho/X^\mu,$$

so that P^λ and X^λ are cogredient.]

VII. 5. Contragredience with linear relation.— (i) The simplest case of contragredience is that in which the contragredient sets are connected by a linear relation.

(*a*) Suppose that, with the notation of § 3, the relation between A_λ and X^λ is

$$A_\mu = a_{\mu\nu} X^\nu.$$

Then, if Q denotes the inner product (cross-product) in (VII. 3. 2),

$$Q = A_\mu X^\mu = a_{\mu\nu} X^\mu X^\nu = a_{11} X^1 X^1 + (a_{12} + a_{21}) X^1 X^2 \\ + a_{22} X^2 X^2 + (a_{13} + a_{31}) X^1 X^3 + \ldots + a_{mm} X^m X^m.$$

Thus Q is a quadratic in $X^1, X^2, X^3, \ldots X^m$, i.e. in X^μ. Also, since each of the four sets X^μ, Y_μ, A_μ, B^μ is a linear function of each of the others, Q can be expressed in a good many other ways, e.g. as a quadratic in B^μ, or in the form $k^{\mu\rho} A_\mu Y_\rho$ or $a^{\mu\nu} A_\mu A_\nu$.

(*b*) If the relation between A_μ and X^μ is symmetrical, i.e. if $a_{\mu\rho} = a_{\rho\mu}$, it can be shown that we shall, in addition to (VII. 3. 1), have the further relations

$$\frac{A_\mu}{Y_\rho} = \frac{B^\rho}{X^\mu}, \quad \frac{Y_\rho}{A_\mu} = \frac{X^\mu}{B^\rho}.$$

(ii) Conversely, suppose that we are dealing with a set

of quantities or coordinates $X^\mu \equiv (X^1 X^2 \ldots X^m)$, and that we come across an expression

$$Q \equiv a_{\mu\nu} X^\mu X^\nu,$$

where $a_{\mu\nu} = a_{\nu\mu}$. Then we may construct a new set given by

$$X_\mu \equiv a_{\mu\nu} X^\nu,$$

and we shall have

$$Q = X^\mu X_\mu.$$

Now suppose we change the system of coordinates linearly or replace the X's by some linear functions of them. The a's will normally have some definite meaning; and this meaning, though not their actual values, will remain unchanged when the X's are changed. Suppose that, when X^μ becomes Y_μ, X_μ—as based on this meaning of the a's—becomes Y^μ. Then we shall have

$$Y_\mu Y^\mu = X^\mu X_\mu,$$

and also

$$\frac{Y_\mu}{X^\nu} = \frac{X_\nu}{Y^\mu} = \frac{Q}{X^\nu Y^\mu}, \text{ etc.}$$

VII. 6. **Ratios of sets generally.**—(i) The word *ratio* has so far only been used in reference to single sets connected by a linear relation; if the relation between Y_ρ and X^μ is of the form $Y_\rho = d_{\mu\rho} X^\mu$, we call $d_{\mu\rho}$ the ratio of Y_ρ to X^μ, and we call $d^{\mu\rho}$ the reciprocal of this ratio. We can extend the use of the word to sets other than single sets.

(ii) We have already had examples of a ratio which involves a scalar. Thus in § 2 we had relations $Z = h_\lambda X^\lambda$, $C = h_\lambda A^\lambda$, and we said that

$$\frac{C}{A^\lambda} = h_\lambda = \frac{Z}{X^\lambda}.$$

Here we can quite well call h_λ the ratio of C to A^λ or of Z to X^λ;

i.e. it is the ratio of a scalar to a single set. Similarly in § 5 (ii) the statement

$$\frac{Y_\mu}{X^\nu} = \frac{Q}{X^\nu Y^\mu}$$

may be regarded as a statement that the ratio of Q to $X^\nu\, Y^\mu$ is equal to that of Y_μ to X^ν. In each case the ratio of one set to another is the set by which the latter has to be multiplied in order to obtain the former.

(iii) The more important case is that of the ratio of a variable set of any rank to another variable set of the same rank. If $\mathfrak{A}$ is a variable set of any rank, and $\mathfrak{B}$ is a set which is a function of $\mathfrak{A}$ and is of the same rank, and if the relation between $\mathfrak{B}$ and $\mathfrak{A}$ is of the form

$$\mathfrak{B} = \mathfrak{p}\mathfrak{A},$$

where $\mathfrak{p}$ is a constant set whose symbol contains all the suffixes occurring in $\mathfrak{A}$ and $\mathfrak{B}$, then we can call $\mathfrak{p}$ the ratio of $\mathfrak{B}$ to $\mathfrak{A}$, and denote this ratio by $\mathfrak{B}/\mathfrak{A}$.

(iv) In these cases we can continue to speak of the relation of $\mathfrak{B}$ to $\mathfrak{A}$ as linear. Also, by solving the equations, we find that there is a relation of the form

$$\mathfrak{A} = \mathfrak{p}'\mathfrak{B},$$

so that, if $\mathfrak{B}$ is a linear function of $\mathfrak{A}$, then $\mathfrak{A}$ is a linear function of $\mathfrak{B}$. Here, $\mathfrak{p}$ being the ratio of $\mathfrak{B}$ to $\mathfrak{A}$, $\mathfrak{p}'$ is the ratio of $\mathfrak{A}$ to $\mathfrak{B}$, and we can call each ratio the reciprocal of the other.

(v) As an example, suppose that $\mathfrak{A}$ and $\mathfrak{B}$ are of rank 3, and that $\mathfrak{p}$ is the product of three double sets, each of which has inner multiplication with $\mathfrak{A}$. Then the relation might be of the form

$$B^\tau_{\rho\sigma} = a_{\lambda\rho} b_{\mu\sigma} c^{\nu\tau} A^{\lambda\mu}_\nu.$$

It is easy to show that in this case

$$A_{\nu}^{\lambda\mu} = a^{\lambda\rho} b^{\mu\sigma} c_{\nu\tau} B_{\rho\sigma}^{\tau}.$$

Thus the reciprocal of the product of the three double sets is the product of their reciprocals.

VII. 7. Related sets of higher rank.—With the preceding explanation, there is no difficulty in extending the ideas of equality of ratios, and of related systems of sets, to sets of higher rank.

Suppose, for instance, that we have a set $\mathfrak{A}$ which is a function of three single sets, and a set $\mathfrak{B}$ which is a function of three other single sets. If the three latter sets were functions of the three former, $\mathfrak{B}$ would be a function of $\mathfrak{A}$. The cases analogous to those considered in § 4 would be the cases in which there were linear relations between corresponding sets. Suppose that $X_\lambda, Y_\lambda, Z_\lambda$ are linear functions of U^λ, V^λ, W^λ respectively, e.g.

$$X_\rho = a_{\lambda\rho} U^\lambda, \quad Y_\sigma = b_{\mu\sigma} V^\mu, \quad Z_\tau = c_{\nu\tau} W^\nu,$$

that $\mathfrak{A}$ is a certain function (in the most general sense) of U^λ, V^λ, W^λ, and that $\mathfrak{B}$ is the same function of X_λ, Y_λ, Z_λ. Then $\mathfrak{B}$ is some function of $\mathfrak{A}$. If we suppose that, when the values of $a_{\lambda\rho}, b_{\mu\sigma}, c_{\nu\tau}$ are made to vary, $\mathfrak{B}$ is always a linear function of $\mathfrak{A}$, and the ratio of $\mathfrak{B}$ to $\mathfrak{A}$ is always compounded of the ratios of X_λ to U^λ, of Y_λ to V^λ, and of Z_λ to W^λ, each taken directly or reciprocally, we get an extension of the cases considered in § 4. Thus we might have

$$B_{\rho\sigma}^{\tau} = a_{\lambda\rho} b_{\mu\sigma} c^{\nu\tau} A_{\nu}^{\lambda\mu},$$

so that

$$\frac{B_{\rho\sigma}^{\tau}}{A_{\nu}^{\lambda\mu}} = a_{\lambda\rho} b_{\mu\sigma} c^{\nu\tau} = \frac{X_\rho}{U^\lambda} \cdot \frac{Y_\sigma}{V^\mu} \cdot \frac{W^\nu}{Z_\tau};$$

each of these latter expressions representing a set of rank 6.

In this particular case we can say that $A_{\nu}^{\lambda\mu}$ is *cogredient* as regards U^{λ} and V^{μ} and *contragredient* as regards W^{ν}; or we can say that it *varies directly* as regards U^{λ} and V^{μ} and *reciprocally* as regards W^{ν}, or that it is *directly proportional* to U^{λ} and V^{μ} and *reciprocally proportional* to W^{ν}.

VIII. DIFFERENTIAL RELATIONS OF SETS

VIII. 1. Derivative of a set.—We have now to consider the cases in which two sets vary together continuously, so that there can be a derivative (differential coefficient) of one with regard to the other, this latter being a single set. The derivative will in all cases be a partial one, since the elements of the single set vary independently.

(i) The simplest case is that of a scalar linear function

$$Z = h_\lambda X^\lambda = h_1 X^1 + h_2 X^2 + \ldots + h_m X^m.$$

Here

$$\frac{\partial Z}{\partial X^p} = h_p.$$

Giving p all values 1 to m, we can write this

$$\frac{\partial Z}{\partial X^\lambda} = h_\lambda = \frac{Z}{X^\lambda};$$

and we can regard h_λ as the derivative of the set Z (which is of rank 0) with regard to the set X^λ.

(ii) Similarly, if

$$Y_\rho = d_{\mu\rho} X^\mu,$$

then *

$$\frac{\partial Y_\rho}{\partial X^\mu} = d_{\mu\rho} = \frac{Y_\rho}{X^\mu}.$$

* It should, however, be noticed that in this statement the sign '=' has not the same meaning in the two places in which it is used. When we say that $\partial Y_\rho/\partial X^\mu = d_{\mu\rho}$, we mean that $\partial Y_r/\partial X^q = d_{qr}$ for all values of q and r: but, when we say that $Y_\rho/X^\mu = d_{\mu\rho}$, we mean that $Y_r = d_{\mu r} X^\mu$ for all values of r.

(iii) A particular case of (ii) is where $Y_\lambda = X^\lambda$. Here $\partial X^r/\partial X^q$ is $= 1$ or 0 according as r is $= q$ or $\neq q$, since the X's vary independently. Hence (see (VII. 2. 3))

$$\frac{\partial X^\rho}{\partial X^\mu} = \frac{X^\rho}{X^\mu} = |^\rho_\mu = |^\mu_\rho = \frac{X^\mu}{X^\rho} = \frac{\partial X^\mu}{\partial X^\rho}. \qquad \textbf{(VIII. 1. 1)}$$

(iv) Taking also sets of higher rank, and not limiting ourselves to linear functions, we see that the derivative of a set with regard to a single set X^λ is a set of rank higher by 1 than that of the original set.

VIII. 2. Derivative of sum or product.—(i) The derivatives of sums and products of sets follow the ordinary laws of derivatives of sums and products; e.g.

$$\frac{\partial(\mathfrak{B}+\mathfrak{C})}{\partial A^\lambda} = \frac{\partial \mathfrak{B}}{\partial A^\lambda} + \frac{\partial \mathfrak{C}}{\partial A^\lambda}, \qquad \textbf{(VIII. 2. 1)}$$

$$\frac{\partial(\mathfrak{B}\mathfrak{C})}{\partial A^\lambda} = \frac{\partial \mathfrak{B}}{\partial A^\lambda}\mathfrak{C} + \mathfrak{B}\frac{\partial \mathfrak{C}}{\partial A^\lambda}. \qquad \textbf{(VIII. 2. 2)}$$

(ii) As a particular case of this last result, take the scalar quadratic form considered in VII. 5.

$$Q \equiv a_{\mu\nu} X^\mu X^\nu$$

Here, taking (VIII. 1. 1) into account, we have

$$\frac{\partial Q}{\partial X^\rho} = a_{\mu\nu} X^\nu + (a_{\mu\nu} X^\mu)\frac{\partial X^\nu}{\partial X^\rho} = a_{\mu\nu} X^\nu + a_{\rho\nu} X^\rho |^\nu_\mu =$$
$$a_{\mu\nu} X^\nu + a_{\rho\mu} X^\rho = (a_{\mu\nu} + a_{\nu\mu}) X^\nu.$$

This can be verified by expressing Q in terms of the X's and finding the partial derivative with regard to X^q in the usual way.

(iii) For an application of this, suppose that

$$a_{\mu\nu} X^\mu X^\nu = b_{\mu\nu} X^\mu X^\nu$$

for all values of the X's. By taking adjoining values of

X^{μ}, we get the result which is expressed by differentiation, namely
$$2a_{\mu\nu}X^{\nu} = 2b_{\mu\nu}X^{\nu}.$$
Differentiating again, or equating coefficients, we find that
$$a_{\mu\nu} = b_{\mu\nu}.$$

VIII. 3. Derivative of function of a set.—(i) If z is a function of y, and y is a function of x, then we know that
$$\frac{dz}{dy}\frac{dy}{dx} = \frac{dz}{dx};$$
and, total and partial differential coefficients being in this case identical,
$$\frac{\partial z}{\partial y}\frac{\partial y}{\partial x} = \frac{\partial z}{\partial x}.$$

(ii) Now suppose that B_{μ} is a function of A^{λ}, and C^{ρ} is a function of B_{μ}. Then, p and r being values of λ and ρ respectively, we know that
$$\frac{\partial C^r}{\partial A^p} = \frac{\partial C^r}{\partial B_1}\frac{\partial B_1}{\partial A^p} + \frac{\partial C^r}{\partial B_2}\frac{\partial B_2}{\partial A^p} + \ldots + \frac{\partial C^r}{\partial B_m}\frac{\partial B_m}{\partial A^p} = \frac{\partial C^r}{\partial B_{\mu}}\frac{\partial B_{\mu}}{\partial A^p}.$$
Hence, giving p and r all their values,
$$\frac{\partial C^{\rho}}{\partial A^{\lambda}} = \frac{\partial C^{\rho}}{\partial B_{\mu}}\frac{\partial B_{\mu}}{\partial A^{\lambda}}. \qquad \textbf{(VIII. 3. 1)}$$

The argument applies to sets of higher rank. If, e.g., we are dealing with $C^{\rho\sigma\cdots}$, where the $\sigma\ldots$ relates to aspects independent of B_{μ}, then
$$\frac{\partial C^{\rho\sigma\cdots}}{\partial A^{\lambda}} = \frac{\partial C^{\rho\sigma\cdots}}{\partial B_{\mu}}\frac{\partial B_{\mu}}{\partial A^{\lambda}}. \qquad \textbf{(VIII. 3. 2)}$$

(iii) As a particular case (see (VIII. 1. 1)),
$$\frac{\partial C^{\rho}}{\partial B^{\mu}}\frac{\partial B^{\mu}}{\partial C^{\sigma}} = \frac{\partial C^{\rho}}{\partial C^{\sigma}} = 1^{\rho}_{\sigma}. \qquad \textbf{(VIII. 3. 3)}$$

(iv) Suppose that the relation between B_λ and A^λ is linear, say

$$B_\mu = p_{\lambda\mu} A^\lambda, \quad A^\lambda = p^{\lambda\mu} B_\mu.$$

Then, replacing C^r or C^ρ by $\mathfrak{C}$, we have

$$\frac{\partial \mathfrak{C}}{\partial B_\mu} = \frac{\partial \mathfrak{C}}{\partial A^\lambda}\frac{\partial A^\lambda}{\partial B_\mu} = \frac{\partial \mathfrak{C}}{\partial A^\lambda} p^{\lambda\mu}.$$

Hence

$$\frac{\partial \mathfrak{C}}{\partial B_\mu} \Big/ \frac{\partial \mathfrak{C}}{\partial A^\lambda} = p^{\lambda\mu} = A^\lambda / B_\mu.$$

Thus A^λ and $\partial \mathfrak{C}/\partial A^\lambda$ are contragredient. We can express this by saying that A^λ and the operator $\partial/\partial A^\lambda$ are contragredient.

(v) The determinant of $\partial B_\mu / \partial A^\lambda$ is the Jacobian of B_μ with regard to A^λ; i.e.

$$\frac{\partial(B_1, B_2, \dots B_m)}{\partial(A^1, A^2, \dots A^m)} = \left| \frac{\partial B_r}{\partial A^q} \right|. \qquad \textbf{(VIII. 3. 4)}$$

From (VIII. 3. 1), taken with (V. 10. 5), we have the ordinary formula for the product of two Jacobians:

$$\left| \frac{\partial B_r}{\partial A^q} \right| \times \left| \frac{\partial C^r}{\partial B_q} \right| = \left| \frac{\partial B_\mu}{\partial A^q}\frac{\partial C^r}{\partial B_\mu} \right| = \left| \frac{\partial C^r}{\partial A^q} \right|. \qquad \textbf{(VIII. 3. 5)}$$

VIII. 4. **Transformation of quadratic form to sum of squares.**—(i) For an example of a Jacobian, take the case in which a quadratic form is to be expressed as the sum of the squares of linear functions of the variables. Let the quadratic form be

$$Q \equiv a_{\mu\rho} X^\mu X^\rho, \qquad (1)$$

where

$$a_{\mu\rho} = a_{\rho\mu}.$$

Let

$$Y_\sigma \equiv b_{\mu\sigma} X^\mu \qquad (2)$$

be a set of linear functions of the X's, so that

$$X^{\mu} = b^{\mu\sigma} Y_{\sigma}. \tag{3}$$

Suppose that the b's are chosen so that (1) gives

$$Q = Y_1 Y_1 + Y_2 Y_2 + \ldots + Y_m Y_m = Y_{\sigma} Y_{\sigma}. \tag{4}$$

Then, by substitution from (2),

$$Q = b_{\mu\sigma} X^{\mu} . b_{\rho\sigma} X^{\rho} = b_{\mu\sigma} b_{\rho\sigma} X^{\mu} X^{\rho}.$$

Hence, by comparison with (1) (see § 2 (iii)),

$$a_{\mu\rho} = b_{\mu\sigma} b_{\rho\sigma}. \tag{5}$$

The Jacobian which we should usually want to find is that of X^{μ} with regard to Y_{σ}, i.e.

$$J \equiv \frac{\partial(X^1, X^2, \ldots X^m)}{\partial(Y_1, Y_2, \ldots Y_m)} = \left| \frac{\partial X^q}{\partial Y_r} \right|.$$

By (3),

$$\frac{\partial X^{\mu}}{\partial Y_{\sigma}} = b^{\mu\sigma};$$

and therefore

$$J = |b^{qr}|. \tag{6}$$

But (5) gives

$$a^{\mu\rho} = b^{\mu\sigma} b^{\rho\sigma},$$

and therefore

$$|a^{qr}| = |b^{q\sigma} b^{r\sigma}| = |b^{qr}| \times |b^{rq}| = JJ.$$

Hence, combining this with (6),

$$J = |b^{qr}| = \{|a^{qr}|\}^{\frac{1}{2}} = 1/\{|a_{qr}|\}^{\frac{1}{2}}. \tag{7}$$

Similarly the Jacobian of Y_{σ} with regard to X^{μ} is

$$J' \equiv \frac{\partial(Y_1, Y_2, \ldots Y_m)}{\partial(X^1, X^2, \ldots X^m)} = \frac{1}{J} = |b_{qr}| = \{|a_{qr}|\}^{\frac{1}{2}}$$
$$= 1/\{|a^{qr}|\}^{\frac{1}{2}}. \tag{8}$$

There are a good many different ways of expressing Q as in (4), but they all give the same two Jacobians.

(ii) Hence we easily obtain the value of the multiple integral

$$D \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}Q}\, dX^1 dX^2 \dots dX^m. \tag{9}$$

Using $\left(\int\right)^m$ to denote $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \dots (m \text{ times})$, we have

$$\begin{aligned} D &= \left(\int\right)^m e^{-\frac{1}{2}Q} J\, dY_1 dY_2 \dots dY_m \\ &= J \int_{-\infty}^{\infty} e^{-\frac{1}{2}Y_1 Y_1} dY_1 \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}Y_m Y_m} dY_m \\ &= \{|\, a^{qr} \,|\}^{\frac{1}{2}} (2\pi)^{\frac{1}{2}m}. \end{aligned} \tag{10}$$

(iii) We shall require, in the next chapter, the value of

$$N/D,$$

where

$$N \equiv \left(\int\right)^m \tfrac{1}{2} X^q \frac{\partial Q}{\partial X^r} e^{-\frac{1}{2}Q}\, dX^1 dX^2 \dots dX^m \tag{11}$$

and D is as above. We find the value of N, as we have found that of D, by expressing everything in terms of Y's. By (2) and (4), and (VIII. 3. 1) and § 2 (ii),

$$\frac{\partial Q}{\partial X^r} = \frac{\partial Y_\tau}{\partial X^r}\frac{\partial Q}{\partial Y_\tau} = 2 b_{r\tau} Y_\tau;$$

and therefore, by (3),

$$\tfrac{1}{2} X^q \frac{\partial Q}{\partial X^r} = b^{q\sigma} b_{r\tau} Y_\sigma Y_\tau,$$

$$\begin{aligned} N &= b^{q\sigma} b_{r\tau} \left(\int\right)^m Y_\sigma Y_\tau e^{-\frac{1}{2}Q} dX^1 dX^2 \dots dX^m \\ &= J b^{q\sigma} b_{r\tau} \left(\int\right)^m Y_\sigma Y_\tau e^{-\frac{1}{2}Q} dY_1 dY_2 \dots dY_m. \end{aligned} \tag{12}$$

Let us write

$$M_t \equiv J b^{q\sigma} \left(\int\right)^m Y_\sigma Y_t e^{-\frac{1}{2}Q} dY_1 dY_2 \dots dY_m, \tag{13}$$

so that

$$N = b_{r\tau} M_\tau. \tag{14}$$

Then M_t consists of m terms due to the m values of σ. Since Q is the sum of the squares of the Y's, the only term which counts in the integration is that for which $\sigma = t$. Also we know that

$$\int_{-\infty}^{\infty} Y_t Y_t e^{-\frac{1}{2}Y_t Y_t} dY_t = \int_{-\infty}^{\infty} e^{-\frac{1}{2}Y_t Y_t} dY_t.$$

Hence it follows that

$$M_t = J b^{qt} \left(\int\right)^m e^{-\frac{1}{2}Q} dY_1 dY_2 \dots dY_m$$
$$= b^{qt} D;$$

and therefore, by (14),

$$N/D = b_{r\tau} b^{q\tau}$$
$$= |_r^q. \tag{15}$$

IX. EXAMPLES FROM THE THEORY OF STATISTICS *

IX. 1. **Preliminary.**—(i) The special feature of a statistical set

$$X_\lambda \equiv (X_1\, X_2\, X_3 \dots X_m),$$

of the kind which we have to consider in this chapter, is that each X has one only of a very large number of actual or possible values, which together constitute the field from which the X is drawn; and the fundamental facts with which we are concerned are the relative frequencies of occurrence of the various possible combinations formed by taking an X from each of the m fields. Thus the X's are variables, and the expression for the relative frequency of joint occurrence of a particular set $(X_1\, X_2 \dots X_m)$ involves the X's of the set, with certain constants. In a large class of cases the constants depend on the mean values of the X's and the mean squares and products of their deviations from their respective means. It is to these cases that the new notation is specially applicable. It may be that some of the X's are drawn from the same field; we shall proceed as if the fields were all different, but this does not affect the validity of the reasoning.

(ii) We shall only consider two kinds of cases.

(*a*) The first kind of case is where our statistical information relates to a large number of individuals, and $X_1\, X_2\, X_3 \dots$ are the measures of specified attributes, such as height, head-length, chest-expansion, intelligence, etc.,

* I have dealt with the problems of this chapter in as general terms as possible. The explanations in small print may help to show the statistical student the way in which the problems actually arise.

of any individual. Here the 'variability' of any X has reference to the different values that it takes for different individuals. The frequency of joint occurrence of particular values of $X_1 X_2 X_3 \ldots$ may be a complicated function of these values and of certain constants which are to be determined.

(*b*) The other kind of case is that in which the question is one of 'graduation' or 'fitting'. Here $X_1 X_2 X_3 \ldots$ are observed values of different quantities (e.g. rates of mortality at successive ages), or are the results of observation of one quantity at different times or by different observers. The 'variability' of any particular X lies in the fact that the observed value contains an unknown error; and our treatment is based on the assumption that a relation of a particular kind holds between the true X's.

In cases of this latter kind it should be noticed that the only things which we treat as variables are the errors in the X's. We may take, as the typical case, the observed rates of mortality $X_1 X_2 X_3 \ldots$ at ages $t_1 t_2 t_3 \ldots$ If we denote the true rates by $\xi_1 \xi_2 \xi_3 \ldots$, then the assumption which we make is really an assumption that ξ is a certain function, with constants to be determined, of t. So far as this function is concerned, ξ and t might be called variables. But, for our purpose, they are not variables. We are concerned with the fixed values $t_1 t_2 t_3 \ldots$ and the corresponding fixed, though unknown, values $\xi_1 \xi_2 \xi_3 \ldots$; the real variables are the differences between the observed values $X_1 X_2 X_3 \ldots$ and the true values $\xi_1 \xi_2 \xi_3 \ldots$.

(iii) There are two reasons why the same mathematical methods apply to subjects so different as relativity and statistical theory. One is that the number of X's in a statistical set may be very large: in the second kind of case mentioned in (ii) it may be as many as 20 or 30. The need of a condensed notation is therefore even

greater than for relativity, where the number of dimensions does not exceed 4. The other reason is that, as in the relativity theory, we are, to a certain extent in the first kind of case, and very largely in the second kind of case, concerned with sets constructed from the original sets by means of linear relations.

(iv) We denote the mean product of the deviations of X_q and X_r by $(X_q . X_r)$, or, more briefly, by f_{qr}; i.e.

$$f_{qr} \equiv (X_q . X_r) \equiv \text{mean value of } (X_q - \text{mean } X_q)(X_r - \text{mean } X_r).$$

It must be clearly understood that, though $(X_q . X_r)$ depends on q and r, it relates to the complete fields from which X_q and X_r are drawn, and, for any particular values of q and r, is not a variable, like X_q and X_r, but a constant.

(v) Although we have defined $(X_q . X_r)$ as a mean value, our dealings with it ultimately depend on the algebraical laws which it follows. These are, first that it satisfies the ordinary laws of multiplication of two expressions X_q and X_r, i.e. that, c being a constant as regards the X's,

$$(X_q . X_r) = (X_r . X_q), \quad (X_q . (X_r + X_s)) = (X_q . X_r) + (X_q . X_s), \quad (X_q . cX_r) = c(X_q . X_r); \qquad \textbf{(IX. 1. 1)}$$

and next that, if u is any linear function of the X's, then $(u . u)$ is positive unless $u = 0$, i.e. that

$$(u . u) > 0 \quad \text{if} \quad u \neq 0. \qquad \textbf{(IX. 1. 2)}$$

It is clear that $(u . u) = 0$ if $u = 0$; for (IX. 1. 1) gives, by putting $c = 0$,

$$(X_q . 0) = 0, \qquad \textbf{(IX. 1. 3)}$$

whence $(u . 0) = 0$ follows by means of (IX. 1. 1) (cf. (vi) below). These are the only properties we shall use; and our results will therefore be true for any meaning of $(X_q . X_r)$ that satisfies these laws, provided, of course, that

$(X_q . X_r)$ is a constant as regards all the X's. And, conversely, we shall only be dealing with sets for which $(X_q . X_r)$ has a meaning and a value for each value of q with each value of r, and satisfies these laws. It will be assumed that the values of $(X_q . X_r)$ are known; or, at any rate, that our results are final when expressed in terms of these values. Whether we are dealing with mean squares or mean products or not, we can call $(X_q . X_r)$ the (.) of X_q and X_r.

In the first kind of case mentioned in (ii) $(X_q . X_q)$ would usually be the mean square of deviation of X_q from its mean, i. e. would be the square of the standard deviation; and $(X_q . X_r)$ would be the mean product of deviations of X_q and X_r from their respective means, and would therefore, in the case of normal correlation, be equal to the product of the standard deviations of X_q and X_r multiplied by their coefficient of correlation. In the second kind of case $(X_q . X_q)$ is the mean square of error of X_q, and $(X_q . X_r)$ is the mean product of errors of X_q and X_r.

(vi) It follows from (IX. 1. 1) that

$$(a_\mu X_\mu . X_r) = a_\mu (X_\mu . X_r), \quad (a_\mu X_\mu . b_\rho X_\rho) = a_\mu b_\rho (X_\mu . X_\rho). \qquad \textbf{(IX. 1. 4)}$$

(vii) If X_λ is a set of the kind considered in this section then so also is any other set which is a linear function of X_λ. Suppose, for instance, that

$$Y^\mu \equiv b^{\lambda\mu} X_\lambda .$$

Then, by (IX. 1. 4),

$$(Y^q . Y^r) = (b^{\lambda q} X_\lambda . b^{\nu r} X_\nu) = b^{\lambda q} b^{\nu r} (X_\lambda . X_\nu),$$

which has a definite meaning for each value of q with each value of r, and can be shown to satisfy the laws stated in (v).

(viii) Our results are also subject to the condition that none of the determinants of the double sets we have to deal with are 0; i.e. that $|(X_q . X_r)| \neq 0$, whether the range

of values of q and r is the whole of the range 1 to m or a part of it only.

IX. 2. **Mean-product set.**—(i) The quantities $(X_q \,.\, X_r)$ constitute a symmetrical double set

$$f_{\mu\rho} \equiv f_{\rho\mu} \equiv \begin{Bmatrix} (X_1 \,.\, X_1) & (X_1 \,.\, X_2) & (X_1 \,.\, X_3) \ldots (X_1 \,.\, X_m) \\ (X_1 \,.\, X_2) & (X_2 \,.\, X_2) & (X_2 \,.\, X_3) \ldots (X_2 \,.\, X_m) \\ \vdots & \vdots & \vdots \qquad\quad \vdots \\ (X_1 \,.\, X_m) & (X_2 \,.\, X_m) & (X_3 \,.\, X_m) \ldots (X_m \,.\, X_m) \end{Bmatrix}. \qquad \textbf{(IX. 2. A)}$$

We call this the **mean-product set.**

(ii) Corresponding to this there is a reciprocal set $f^{\mu\rho} = f^{\rho\mu}$ given by

$$f^{\lambda\rho} f_{\mu\rho} = f^{\lambda\rho} f_{\rho\mu} = f^{\rho\lambda} f_{\mu\rho} = f^{\rho\lambda} f_{\rho\mu} = |^{\lambda}_{\mu}. \qquad \textbf{(IX. 2. 1)}$$

(iii) If $(X_p \,.\, X_s) = 0$, we can for the purpose of this chapter describe X_p and X_s as *statistically independent.* Strictly speaking, this is a loose description, since the complete statistical independence of two variables X_p and X_s would imply a good deal more than that the mean product of their deviations from their respective means should be 0. But we are only concerned, here, with mean squares and mean products.

(iv) The simplest class of cases—from the point of view of algebraical treatment—consists of those cases in which the X's are statistically independent of one another and the mean square of deviation of each X is 1. We can express this by duplicating the set, thus:

$$\begin{matrix} X_1 & X_2 & X_3 \ldots X_m \\ X_1 & X_2 & X_3 \ldots X_m \end{matrix}$$

and saying that the (.) of corresponding elements of the two sets is 1, and that the (.) of elements which do not correspond is 0.

For a set of this kind we have

$$f_{qr} = (X_q \, . \, X_r) = |_r^q,$$

so that the mean-product set is the unit set. It follows that in this class of cases the mean-product set and its reciprocal set are identical.

(v) The next kind of case, in point of simplicity, is that in which the X's are statistically independent of one another but the mean squares of deviation are not all 1. This would be the case, for instance, if the X's were independent observations, not all of the same weight, of a single quantity. For practical purposes a case of this kind can be brought under (iv) by expressing each X in terms of its standard deviation (square root of mean square of deviation) as the unit.

(vi) There is also an important class of cases in which the X's fall into two groups, such that each X in one group is statistically independent of each X in the other group. If, as in VI. 11 (i), we denote the two groups by A_a and A_ϕ, then the property is that

$$(A_a \, . \, A_\phi) = 0.$$

IX. 3. Conjugate sets.—(i) When a set X_λ is not of the simple kind described in § 2 (iv), we shall find it useful to introduce another set X^λ which (1) is a linear function of X_λ and (2) is such that, if we place the sets opposite one another, thus:

$$\begin{matrix} X_1 & X_2 & X_3 \ldots X_m \\ X^1 & X^2 & X^3 \ldots X^m \end{matrix}$$

the (.) of corresponding elements of the two sets is 1, and that of elements which do not correspond is 0. This new set X^λ is said to be **conjugate** to X_λ.

(ii) The second of the above conditions can be written

$$(X^p \, . \, X_q) = |_q^p, \qquad \textbf{(IX. 3. 1)}$$

or

$$(X^\lambda \,.\, X_\mu) = |^\lambda_\mu. \qquad \textbf{(IX. 3. 2)}$$

(iii) Each element X^p of the new set will contain m terms, with m coefficients which have to be determined from the m equations given by

$$(X^p \,.\, X_\mu) = |^p_\mu.$$

There are altogether m^2 equations to determine X^λ. By regrouping these according to the values of μ in $(X^\lambda \,.\, X_\mu)$, we see that if X^λ is conjugate to X_λ then X_λ is conjugate to X^λ. This is in fact evident from the symmetry of (IX. 3. 2).

(iv) To express X^λ in terms of X_λ, or X_λ in terms of X^λ, let us first take W to be any linear function of X_λ, say

$$W = a^\mu X_\mu. \qquad (1)$$

Then we want to find an expression for a^μ.

As we know the value of $(X^\lambda \,.\, X_\mu)$, we take the (.) of W and X^λ. By (IX. 1. 4) and (IX. 3. 2) we find that

$$(W \,.\, X^\lambda) = (a^\mu X_\mu \,.\, X^\lambda) = a^\mu (X_\mu \,.\, X^\lambda) = a^\mu |^\lambda_\mu = a^\lambda,$$

whence

$$a^\mu = (W \,.\, X^\mu).$$

Substituting in (1),

$$W = (W \,.\, X^\mu)\, X_\mu. \qquad \textbf{(IX. 3. 3)}$$

Taking W to be each element of X^λ in turn, we have

$$X^\lambda = (X^\lambda \,.\, X^\mu)\, X_\mu. \qquad \textbf{(IX. 3. 4)}$$

We do not yet know the values of $(X^\lambda \,.\, X^\mu)$. But, if we had started with W as a linear function of X^λ, we should similarly have got

$$W = (W \,.\, X_\mu)\, X^\mu, \qquad \textbf{(IX. 3. 5)}$$

whence

$$X_\lambda = (X_\lambda \,.\, X_\mu)\, X^\mu. \qquad \textbf{(IX. 3. 6)}$$

Writing this in the form

$$X_\lambda = f_{\lambda\mu} X^\mu, \qquad \textbf{(IX. 3. 7)}$$

we have, by reciprocation,

$$X^\mu = f^{\lambda\mu} X_\lambda, \qquad \textbf{(IX. 3. 8)}$$

which gives X^μ in terms of X_λ. Further, comparing (IX. 3. 8) with (IX. 3. 4), we see that

If $(X_\lambda . X_\mu) \equiv f_{\lambda\mu}$, then $(X^\lambda . X^\mu) = f^{\lambda\mu}$; **(IX. 3. 9)**

and, of course, the converse also holds. This result is dependent on the assumption, made in § 1 (viii), that $|f_{qr}|$ is not 0.

(v) We could have obtained (IX. 3. 4) and (IX. 3. 6) in fewer steps by considering the set as a whole instead of element by element. If we assume

$$X^\lambda = a^{\lambda\mu} X_\mu,$$

then we get

$$(X^\lambda . X^\nu) = (a^{\lambda\mu} X_\mu . X^\nu) = a^{\lambda\mu} |_\mu^\nu = a^{\lambda\nu},$$

so that

$$a^{\lambda\mu} = (X^\lambda . X^\mu).$$

This gives (IX. 3. 4); and (IX. 3. 6) can be obtained in the same way.

(vi) We can write (IX. 3. 3) in the form

$$W/X_\mu = (W . X^\mu); \qquad \textbf{(IX. 3. 10)}$$

and similarly from (IX. 3. 5)

$$W/X^\mu = (W . X_\mu). \qquad \textbf{(IX. 3. 11)}$$

(vii) If the set X_λ is of the special kind considered in § 2 (iv), i.e. is such that

$$(X_\lambda . X_\mu) = |_\mu^\lambda,$$

then

$$X_\lambda = (X_\lambda . X_\mu) X^\mu = |_\mu^\lambda X^\mu = X^\lambda,$$

so that the set is identical with its conjugate. The set is said to be **self-conjugate.**

(viii) In a case of the kind mentioned in § 2 (v), where $f_{pq} = 0$ if $q \neq p$, but f_{pp} is not necessarily $= 1$, it may be shown that $f^{pp} = 1/f_{pp}$, and $f^{pq} = 0$ if $q \neq p$, so that $X^p = X_p/f_{pp}$.

(ix) Next consider a case of the kind mentioned in § 2 (vi), where X_λ consists of two portions, the elements in each portion being independent of those in the other portion. As before, we take one portion to consist of the first k elements, and the other of the remaining $m-k$, and we denote the two portions by X_α and X_ϕ. Then the special property is that

$$f_{\alpha\phi} \equiv (X_\alpha \,.\, X_\phi) = 0, \tag{1}$$

whence, as in VI. **11** (iii), it follows that

$$f^{\alpha\phi} = 0. \tag{2}$$

Breaking up the right-hand side of (IX. 3. 7) into two portions, we get, according as λ belongs to the first or to the second portion, the two separate results

$$X_\alpha = f_{\alpha\gamma} X^\gamma, \quad X_\phi = f_{\phi\psi} X^\psi. \tag{3}$$

Similarly, from (IX. 3. 8),

$$X^\alpha = f^{\alpha\gamma} X_\gamma, \quad X^\phi = f^{\phi\psi} X_\psi. \tag{4}$$

In finding $f^{\alpha\gamma}$ and $f^{\phi\psi}$ from $f_{\lambda\mu}$, it is (see VI. **11** (iii)) immaterial whether we take the set $f_{\lambda\mu}$ as a whole or the sets $f_{\alpha\gamma}$ and $f_{\phi\psi}$ separately, so that (4) may equally well be written (see VI. **11** (ii))

$$X^\alpha = (f^{\alpha\gamma})_k X_\gamma, \quad X^\phi = [f^{\phi\psi}]_{m-k} X_\psi. \tag{5}$$

IX. 4. Conjugate sets with linear relations.—(i) Let Y_ρ be any set which is a linear function of X_μ, and therefore of X^μ; and let Y^ρ be its conjugate set. Then, taking W in (IX. 3. 10) to be each element of Y_ρ in turn, we have

$$Y_\rho/X_\mu = (Y_\rho \,.\, X^\mu). \tag{IX. 4. 1}$$

Similarly

$$Y_\rho/X^\mu = (Y_\rho \,.\, X_\mu), \quad Y^\rho/X_\mu = (Y^\rho \,.\, X^\mu), \quad Y^\rho/X^\mu = (Y^\rho \,.\, X_\mu). \tag{IX. 4. 2}$$

(ii) By combining ratios, we get such results as

$$(E_\sigma \,.\, X^\mu) = E_\sigma/X_\mu = E_\sigma/Y_\rho \,.\, Y_\rho/X_\mu = (E_\sigma \,.\, Y^\rho)\,(Y_\rho \,.\, X^\mu). \qquad \textbf{(IX. 4. 3)}$$

(iii) To find Y^ρ, suppose that

$$Y_\rho = b_{\rho\mu}\, X_\mu. \qquad (1)$$

Let the conjugate set be

$$Y^\rho = k_{\rho\mu}\, X^\mu.$$

Then

$$|^\rho_\sigma = (Y^\rho \,.\, Y_\sigma) = (k_{\rho\mu} X^\mu \,.\, b_{\sigma\nu} X_\nu) = |^\mu_\nu k_{\rho\mu} b_{\sigma\nu} = k_{\rho\mu} b_{\sigma\mu};$$

whence, by reciprocation,

$$k_{\rho\mu} = b^{\sigma\mu} |^\rho_\sigma = b^{\rho\mu}.$$

Thus the conjugate set is

$$Y^\rho = b^{\rho\mu}\, X^\mu. \qquad (2)$$

(iv) Similarly, if

$$Y_\epsilon = b_{\epsilon\nu}\, c_{\nu\rho}\, d_{\rho\sigma}\, X_\sigma,$$

then

$$Y^\epsilon = b^{\epsilon\nu} c^{\nu\rho} d^{\rho\sigma} X^\sigma,$$

etc.; in other words, the conjugate of an inner product is the inner product of the $\left\{\begin{matrix}\text{reciprocals}\\ \text{conjugate}\end{matrix}\right\}$ of the factors. [Let us write

$$C^\epsilon \equiv b^{\epsilon\eta} c^{\eta\theta} d^{\theta\lambda} X^\lambda.$$

Then

$$(C^\epsilon \,.\, Y_\varsigma) = b^{\epsilon\eta} b_{\varsigma\nu} c^{\eta\theta} c_{\nu\rho} d^{\theta\lambda} d_{\rho\sigma} (X^\lambda \,.\, X_\sigma)$$
$$= |^\epsilon_\varsigma,$$

since $(X^\lambda \,.\, X_\sigma) = |^\lambda_\sigma$. Hence

$$C^\epsilon = Y^\epsilon.$$

We might, alternatively, have deduced this from (iii) by means of (VI. 10. 2).]

(v) From (IX. 4. 1)

$$Y_\rho/X_\mu = (Y_\rho \,.\, X^\mu) = (X^\mu \,.\, Y_\rho) = X^\mu/Y^\rho;$$

and therefore, by VII. 3 (iv),

$$Y_\rho\, Y^\rho = X_\mu\, X^\mu. \qquad \textbf{(IX. 4. 4)}$$

Thus conjugate sets are contragredient (VII. 4); and the

inner product of a set and its conjugate is the same for all linearly related sets. If we denote this inner product by Q, then the sets X_μ, X^μ, Y_μ, Y^μ are connected by four relations of the form

$$\frac{X_\mu}{Y^\rho} = \frac{Y_\rho}{X^\mu} = (X_\mu \,.\, Y_\rho) = \frac{Q}{X^\mu Y^\rho}. \qquad \textbf{(IX. 4. 5)}$$

We can express Q in such forms as

$$Q = X_\mu X^\mu = (X_\lambda \,.\, X_\mu) X^\lambda X^\mu = (X^\lambda \,.\, X^\mu) X_\lambda X_\mu,$$

the last of which, when written out in full, is

$$f^{11} X_1 X_1 + 2 f^{12} X_1 X_2 + f^{22} X_2 X_2 + 2 f^{13} X_1 X_3 + \ldots + f^{mm} X_m X_m,$$

or in more general forms such as

$$Q = (C^\lambda \,.\, D^\mu)\, C_\lambda D_\mu,$$

where C_λ and D_λ are any linear functions of X_λ. It must be remembered that the invariability of Q only applies for the particular values $X_1, X_2, \ldots X_m$. If there were a different set of X's there would (in general) be a different Q.

IX. 5. **The frequency-quadratic.**—(i) In most of the cases we are considering, whether of the first or of the second kind mentioned in § 1 (ii), the frequency of joint occurrence of values lying within limits

$$X_1 \pm \tfrac{1}{2} d X_1, \; X_2 \pm \tfrac{1}{2} d X_2, \ldots X_m \pm \tfrac{1}{2} d X_m$$

is proportional to

$$e^{-\frac{1}{2}P} dX_1 dX_2 \ldots dX_m,$$

where, if $x_1, x_2, \ldots x_m$ are the deviations of $X_1, X_2 \ldots X_m$ from their respective means, P is of the form

$$P \equiv a^{11} x_1 x_1 + 2 a^{12} x_1 x_2 + a^{22} x_2 x_2 + 2 a^{13} x_1 x_3 + \ldots + a^{mm} x_m x_m. \quad (1)$$

In our notation this becomes

$$P \equiv a^{\lambda\mu} x_\lambda x_\mu\,; \qquad (2)$$

where

$$a^{\lambda\mu} = a^{\mu\lambda}. \qquad (3)$$

(ii) Let us write

$$E^\lambda \equiv a^{\lambda\mu} x_\mu \equiv a^{\lambda 1} x_1 + a^{\lambda 2} x_2 + \ldots + a^{\lambda m} x_m$$
$$= \tfrac{1}{2} \partial P / \partial x_\lambda. \tag{4}$$

Then it can be shown (see VIII. 4 (iii)) that

$$\text{mean value of } E^p x_q = \begin{Bmatrix} 1 \text{ if } q = p \\ 0 \text{ if } q \neq p \end{Bmatrix}. \tag{5}$$

Taking $(B_p \,.\, C_q)$ to mean, for these cases, the mean product of B_p and C_q, it follows from (5) that E^λ is conjugate to x_λ, i.e.

$$E^\lambda = x^\lambda. \tag{6}$$

Hence, by (4),

$$x^\lambda = a^{\lambda\mu} x_\mu,$$

and therefore, by (IX. 3. 10), or by comparison with (IX. 3. 4),

$$a^{\lambda\mu} = x^\lambda / x_\mu = (x^\lambda \,.\, x^\mu). \tag{7}$$

(iii) Thus the a's in the expression for P given in (1) are the mean squares and mean products of the elements of the conjugate set. Similarly, if we expressed P in terms of the conjugate set, the coefficients would be the mean squares and mean products of the elements of the original set; i.e.

$$\left.\begin{aligned} P &= a^{\lambda\mu} x_\lambda x_\mu = x_\lambda x^\lambda = a_{\lambda\mu} x^\lambda x^\mu \\ &= a_{11} x^1 x^1 + 2 a_{12} x^1 x^2 + a_{22} x^2 x^2 + 2 a_{13} x^1 x^3 + \ldots + a_{mm} x^m x^m, \end{aligned}\right\} \tag{8}$$

where

$$a_{\lambda\mu} = (x_\lambda \,.\, x_\mu). \tag{9}$$

(iv) Take, for example, the case of two quantities X_1, X_2, whose standard deviations and coefficient of correlation are c_1, c_2, and r. Then it is well known that

$$P = \left(\frac{x_1 x_1}{c_1 c_1} - 2 \frac{r x_1 x_2}{c_1 c_2} + \frac{x_2 x_2}{c_2 c_2} \right) \Big/ (1 - rr).$$

This gives, for the members of the conjugate set,

$$x^1 = \left(\frac{x_1}{c_1} - \frac{rx_2}{c_2}\right) \times \frac{1}{c_1(1-rr)}, \quad x^2 = \left(-\frac{rx_1}{c_1} + \frac{x_2}{c_2}\right) \times \frac{1}{c_2(1-rr)}.$$

It is easily verified that the mean square of x^1 is equal to the coefficient of $x_1 x_1$ in P, and so on.

(v) If we express the x's linearly in terms of a new set y_λ, the value of P will remain unaltered. We can put this differently as follows. Suppose that y_λ is any linear function of x_λ. Let $h_{\lambda\mu} \equiv (y_\lambda \cdot y_\mu)$; and let $h^{\lambda\mu}$ be the reciprocal of $h_{\lambda\mu}$. Then, if we write

$$y^\lambda = h^{\lambda\rho} y_\rho,$$

we shall have

$$y^\lambda y_\lambda = P = x^\lambda x_\lambda.$$

(vi) The $(x_q \cdot x_q)$ or $(x_q \cdot x_r)$ which we have so far been considering is the mean square of x_q, or the mean product of x_q and x_r, without regard to the values that each of the other x's may have; i.e. the mean square or mean product taken for all possible values of these other x's according to their relative frequencies. We may also want to know what happens when some of the x's have definite values ascribed to them and are not allowed to vary from these values. In these cases we follow the principle of VI. 11 (ii). Suppose that all the x's after x_k are fixed. Let the x's up to x_k be denoted by x_α or x_β etc. Then our methods apply to the set of order k formed by these x's. The principle, therefore, is as follows. Suppose we want to study the variation of the k quantities x_α when the $m-k$ quantities x_ϕ are fixed. We first construct the mean-product set $f_{\mu\rho}$ $(=a_{\mu\rho})$ for the m quantities x_α and x_ϕ; then construct the reciprocal set $f^{\mu\rho}$ $(=a^{\mu\rho})$, the elements of which are the coefficients in the terms in P; then take out the partial set $f^{\alpha\gamma}$ corresponding to x_α; and then find the set $(f_{\alpha\gamma})_k$ which is the reciprocal of $f^{\alpha\gamma}$. The result is the mean-product

set of x_a when x_ϕ is fixed. (This is a well-known theorem, but is usually expressed in terms of determinants.)

(vii) If the partial sets x_a and x_ϕ in (vi) are independent, so that $(x_a . x_\phi) = 0$, the elements of the mean-product set outside the portions corresponding to $(x_a . x_\gamma)$ and $(x_\phi . x_\psi)$ will all be 0, and (see § 3 (ix)) the values of $(f_{a\gamma})_k$ will be the same whether we construct them from the whole set or from the partial set. This is otherwise obvious; for, if the x's of x_a vary independently of those of x_ϕ, they vary in the same way whether the latter are fixed or vary.

IX. 6. Criteria for improved values.—(i) Our next problem, considered in this and the following section, is that of reduction of error, in a case of the second kind mentioned in § 1 (ii). We have a set of quantities

$$D_\lambda \equiv (D_1 \quad D_2 \quad D_3 \ldots D_m)$$

which contain errors; the mean products of error being

$$d_{\lambda\mu} \equiv (D_\lambda . D_\mu). \qquad \textbf{(IX. 6. A)}$$

The whole set D_λ consists of two portions

$$D_a \equiv (D_1 \quad D_2 \ldots D_k), \quad D_\phi \equiv (D_{k+1} \quad D_{k+2} \ldots D_m).$$

All the D's are the results (direct or indirect) of observation; but the true values of the D_ϕ are negligible (within the degree of accuracy to which we are working), and, if U is any one of the D's, or any linear function of them, we can add to it any linear function* of D_ϕ without altering, except to a negligible extent, the true value which it represents. If the sum of U and an indeterminate linear function of D_ϕ is represented by U', the problem of *reduction of error* is to determine this linear function so as to make

* This is, of course, an incomplete statement. We could replace U by any function of U and D_ϕ which would be equal to U if the D_ϕ were all 0. But we are only considering linear functions.

$(U'.U')$ a minimum. The resulting value of U' is called the **improved value** of U, and will be denoted by IU. The elements of D_ϕ are called the **auxiliaries.** We can replace a by $\beta, \gamma, \ldots$, and ϕ by $\chi, \psi, \ldots$, as occasion requires.

(ii) The way in which this problem arises is as follows. We start with a set of observed quantities $X_1, X_2, \ldots X_m$, which correspond to a series of values of some other quantity t at equal or unequal intervals; the X's might, e.g., be rates of mortality at different ages. The X's contain errors; and our fundamental assumption, based on general experience and on inspection of the particular *data*, is that the true values are so nearly of the form (in ordinary notation) $c_0 + c_1 t + c_2 t^2 + \ldots + c_{k-1} t^{k-1}$ that their differences (divided differences if the values of t are at unequal intervals) after the $(k-1)$th are negligible. We may therefore add to each X any linear function of these differences, which are what we are calling $D_{k+1}, D_{k+2}, \ldots D_m$. The problem is to determine the coefficients in this linear function so that the mean square of error of the sum of the X and the linear function shall be a minimum.

(iii) We have first to see what relations hold between the two portions of D_λ and the two portions of its conjugate set when similarly divided. Denoting the conjugate set, as usual, by D^λ, let the two portions be

$$D^a \equiv (D^1 \quad D^2 \ldots D^k), \quad D^\phi \equiv (D^{k+1} \quad D^{k+2} \ldots D^m).$$

Here it is to be observed that D^a is not (in general) the set (order k) conjugate to the set D_a (order k), since each element of it is a linear function of the whole D_λ; and similarly for D^ϕ. Now the condition of conjugacy is

$$(D^p . D_q) = 1^p_q.$$

But, if D^p and D_q belong to non-corresponding portions of the two sets, q cannot be equal to p. Hence we get the relations

$$(D^a . D_\beta) = |^a_\beta \ldots (1), \quad (D^a . D_\phi) = 0 \ldots (2),$$
$$(D^\phi . D_a) = 0 \ldots (3), \quad (D^\phi . D_\chi) = |^\phi_\chi \ldots (4).$$

(iv) First let U be an element of D_ϕ or a linear function of D_ϕ, say $a^\phi D_\phi$. Then its improved value must be $U - U = 0$; i. e.

$$I(a^\phi D_\phi) = 0. \qquad \textbf{(IX. 6. 1)}$$

For this makes $(U' . U') = (0 . 0) = 0$, by (IX. 1. 3); and, by (IX. 1. 2), $(U' . U')$ would be > 0 if U' were not $= 0$. Hence $(U' . U')$ is a minimum when $U' = 0$.

(v) The next most simple case is that in which U is an element of D^a or a linear function of D^a, say

$$U \equiv a_a D^a. \qquad (1)$$

Let the value of U' be

$$U' = U + u,$$

where

$$u \equiv a^\phi D_\phi. \qquad (2)$$

Then, by (IX. 1. 1),

$$(U' . U') = (U . U) + 2(U . u) + (u . u).$$

But, by (1) and (2), and by (2) of (iii) above,

$$(U . u) = a_a a^\phi (D^a . D_\phi) = 0.$$

Hence $(U' . U')$ is a minimum when $(u . u)$ is a minimum; and this, by (IX. 1. 2), is when $u = 0$, so that

$$U' = U.$$

Hence the improved value is the same as the original value; i. e.

$$I(a_a D^a) = a_a D^a. \qquad \textbf{(IX. 6. 2)}$$

(vi) The simplicity of the results obtained in (iv) and (v) suggests that we should in all cases regard U as expressed in terms of D^a and D_ϕ. There is no difficulty about this,

in theory, whatever linear function of the D's U may be. If, for instance, U is given as a linear function of D_λ, then we obtain our result by eliminating D_a (k values) between this formula for U and the k equations which give D^a in terms of D_λ, i.e. in terms of D_a and D_ϕ. Suppose then that

$$U = V + W,$$

where

$$V = b_a D^a, \quad W = b^\phi D_\phi.$$

Then U' is formed from U by adding some linear function of D_ϕ, so that

$$U' = V + W',$$

where $V = b_a\, D^a$ as before, and W' is of the form

$$W' = c^\phi D_\phi.$$

Hence

$$(U'\,.\,U') = (V\,.\,V) + 2(V\,.\,W') + (W'\,.\,W').$$

But

$$(V\,.\,W') = (b_a D^a\,.\,c^\phi D_\phi) = b_a c^\phi (D^a\,.\,D_\phi) = 0,$$

by (2) of (iii). Hence

$$(U'\,.\,U') = (V\,.\,V) + (W'\,.\,W').$$

But this is a minimum when $W' = 0$. Hence

$$I(b_a D^a + b^\phi D_\phi) = b_a D^a. \qquad \textbf{(IX. 6. 3)}$$

In other words, if we express U in terms of D^a and D_ϕ, the improved value of U is found by omitting the part involving D_ϕ.

(vii) Since, by (IX. 6. 3), IU is a linear function of D^a, and, by (2) of (iii), $(D^a\,.\,D_\phi) = 0$, it follows, by (IX. 1. 4), that

$$(IU\,.\,D_\phi) = 0. \qquad \textbf{(IX. 6. 4)}$$

In other words, the (.) of any improved value and each of the auxiliaries is 0.

(viii) It also follows from (IX. 6. 3) that if two quantities differ by a linear function of the auxiliaries they have the same improved value.

(ix) By taking U in (vi) to be each member, in turn, of a set B_λ of linear functions of the D's expressed in the form

$$B_\lambda \equiv b_{\lambda a} D^a + b^{\lambda\phi} D_\phi,$$

we find that

$$IB_\lambda = b_{\lambda a} D^a.$$

Also

$$I(k^\lambda B_\lambda) = I(k^\lambda b_{\lambda a} D^a + k^\lambda b^{\lambda\phi} D_\phi) = k^\lambda b_{\lambda a} D^a = k^\lambda IB_\lambda; \quad \textbf{(IX. 6. 5)}$$

i. e. the improved value of any linear function of the B's is the same linear function of their improved values.

(x) Altering k^λ in (IX. 6. 5) to k^λ_μ, and writing $C_\mu \equiv k^\lambda_\mu B_\lambda$, we find that

$$\frac{IC_\mu}{IB_\lambda} = \frac{C_\mu}{B_\lambda}; \quad \textbf{(IX. 6. 6)}$$

i. e. the improved values of two linearly connected sets are related in the same way as the original sets; or, more briefly, a set and its improved values are cogredient.

(xi) Since we know that the improved values of D_ϕ are 0, we have really only to determine those of k other quantities. In view of (IX. 6. 5), we can choose these to be any linear functions of D_a that we like, with or without linear functions of D_ϕ added; and similarly we can replace D_ϕ by any linear functions of D_ϕ: provided, in both cases, that none of the mean-product determinants are 0. The functions so chosen can be called D_a and D_ϕ, so that we need only consider the problem of finding ID_a.

(xii) The result stated in (IX. 6. 2) gives us the extension, to the general case in which the errors of the original observations may

have any mean squares and mean products, of the 'method of moments' ordinarily applied to the case of a self-conjugate set (§ 3 (vii)). We take X_λ, as in (ii), to be the original observations, and D_λ to be their differences of successive orders. Then we have found that the improved value of any linear function of D^a is the same as the original value. But, by VII. 3 (v), $D^1, D^2, \ldots D^k$ are successive sums of the elements of X^λ, the set conjugate to X_λ; and the first k moments of X^λ are linear functions of these sums. Hence the improved values of these moments are the same as their original values; and this, by (ix), is the same thing as saying that the moments of the improved values of X^λ are equal to the moments of the original values. Thus the method of moments still applies. But it should be observed that it does not apply to the original set of observations, but to the conjugate set.

As a simple example, suppose that X_λ is a set of independent observations of a single quantity, the mean square of error of X_p being $c_p c_p$. Then (§ 3 (viii)) the conjugate set is

$$\left(\frac{X_1}{c_1 c_1}\ \frac{X_2}{c_2 c_2} \cdots\cdots \frac{X_m}{c_m c_m}\right).$$

As the X's will all have the same improved value, which we will call IX, there is only one moment to be considered, namely, the 0th moment, or sum, of the conjugate set. Hence, equating the sums of original and of improved values,

$$\frac{X_1}{c_1 c_1} + \frac{X_2}{c_2 c_2} + \ldots + \frac{X_m}{c_m c_m} = \frac{IX}{c_1 c_1} + \frac{IX}{c_2 c_2} + \ldots + \frac{IX}{c_m c_m},$$

which gives the familiar result.

IX. 7. Determination of improved values.—

(i) From the results obtained in the preceding section we deduce three methods of finding the improved value of any element of D_a, say D_f.

(1) We can express D_f in terms of D^a and D_ϕ, and then omit the part involving D_ϕ. The result is ID_f.

(2) We can say that ID_f is some linear function of D^a. This linear function has k coefficients to be determined; they are determined by the condition that, if the linear

function is expressed in terms of D_λ, the coefficient of D_f is 1 and those of other elements of D_a are 0. The practical application of this method depends on the circumstances of the particular class of cases.

(3) We can say that ID_f is obtained from D_f by adding a linear function of D_ϕ, which we have called $-b^\phi D_\phi$. This linear function has $m-k$ coefficients to be determined. We have found in (IX. 6. 4) that $(ID_f \,.\, D_\phi)=0$; this gives $m-k$ equations, from which the coefficients in question can be determined. Thus *the necessary and sufficient conditions for ID_f are that it differs from D_f by a linear function of D_ϕ and that $(ID_f \,.\, D_\phi)=0$.*

These three methods are exhibited in (ii), (iv), and (v) below, and a fourth method is given in (vii).

(ii) To apply the first method, let us write

$$D_a = e_{a\beta} D^\beta + e^{a\phi} D_\phi.$$

We do not need $e^{a\phi}$, since the only part of D_a that counts for the improved value is $e_{a\beta} D^\beta$; we therefore get rid of $e^{a\phi}$ at once, by means of something whose $(\,.\, D_\phi)$ is 0. This, by (2) of § 6 (iii), is D^γ. Taking the $(D^\gamma \,.\,)$ of both sides, we have

$$(D^\gamma \,.\, D_a) = e_{a\beta}(D^\gamma \,.\, D^\beta) = e_{a\beta} d^{\beta\gamma}.$$

Also, by (1) of § 6 (iii),

$$(D^\gamma \,.\, D_a) = |^\gamma_a.$$

Hence

$$d^{\beta\gamma} e_{a\beta} = |^\gamma_a.$$

Here a, β, γ relate to the partial set $(D_1\, D_2 \,.\,.\,.\, D_k)$, and the statement is limited to this set. Dealing only with this set, let us denote the reciprocal of $d^{\beta\gamma}$ by $(d_{\beta\gamma})_k$; this, as pointed out in VI. 11 (ii) (cf. § 5 (vi) of the present chapter) is not ordinarily the same thing as $d_{\beta\gamma}$ as

obtained from the whole set of order m. We have then, by reciprocation,

$$e_{\alpha\beta} = (d_{\beta\gamma})_k |_{\alpha}^{\gamma} = (d_{\beta\alpha})_k.$$

Substituting in the expression for D_α, and dropping the $e^{\alpha\phi} D_\phi$ in order to get the improved value, we have

$$ID_\alpha = (d_{\beta\alpha})_k D^\beta. \qquad \textbf{(IX. 7. 1)}$$

(iii) Although, in the above, we have not needed $e^{\alpha\phi}$, we ought to find its value in order to satisfy ourselves that, as has been stated in § 6 (vi), any linear function of D_λ, say $g_\alpha D_\alpha + g_\phi D_\phi$, can be expressed as a linear function of D^α and D_ϕ; to do this, it is only necessary to prove the proposition for D_α, since the formula for $g_\alpha D_\alpha + g_\phi D_\phi$ will follow at once.

We have written

$$D_\alpha = e_{\alpha\beta} D^\beta + e^{\alpha\phi} D_\phi,$$

and have found $e_{\alpha\beta}$. To find $e^{\alpha\phi}$, we must get rid of the first term; so we again use (2) of § 6 (iii), getting

$$(D_\alpha \,.\, D_\chi) = e^{\alpha\phi} (D_\phi \,.\, D_\chi)$$

or

$$d_{\alpha\chi} = e^{\alpha\phi} d_{\phi\chi}.$$

Hence, by reciprocation,

$$e^{\alpha\phi} = [d^{\phi\chi}]_{m-k} d_{\alpha\chi},$$

where $[d^{\phi\chi}]_{m-k}$ is the reciprocal of $d_{\phi\chi}$ obtained from the partial set $(D_{k+1}\, D_{k+2} \ldots D_m)$. The complete expression for D_α is therefore

$$D_\alpha = (d_{\beta\alpha})_k D^\beta + [d^{\phi\chi}]_{m-k} d_{\alpha\chi} D_\phi. \qquad \textbf{(IX. 7. 2)}$$

The existence of $(d_{\beta\alpha})_k$ and $[d^{\phi\chi}]_{m-k}$ is dependent on the assumption that the determinant $| d_{qr} |$ formed for D_α, and the determinant $| d_{qr} |$ formed for D_ϕ, are both $\neq 0$ (see § 1 (viii)).

(iv) To use the second method, we might have proceeded as follows. We write

$$ID_a = e_{a\beta} D^\beta.$$

To find $e_{a\beta}$, we express D^β in terms of D_μ, i.e. of D_γ and D_ψ, by means of (IX. 3. 4), and we have

$$\begin{aligned} ID_a &= e_{a\beta} d^{\beta\mu} D_\mu \\ &= e_{a\beta} d^{\beta\gamma} D_\gamma + e_{a\beta} d^{\beta\psi} D_\psi. \end{aligned}$$

From the condition stated in (2) of (i), it follows that

$$e_{a\beta} d^{\beta\gamma} = |_a^\gamma,$$

and therefore, by reciprocation,

$$e_{a\beta} = (d_{\beta\gamma})_k |_a^\gamma = (d_{\beta a})_k.$$

Hence we get the same result as before, namely,

$$ID_a = (d_{\beta a})_k D^\beta.$$

(v) For the third method, we write

$$ID_a = D_a - e^{a\phi} D_\phi,$$

and we have to find $e^{a\phi}$. The condition stated in (3) of (i), namely,

$$(ID_a \,.\, D_\chi) = 0,$$

gives

$$d_{a\chi} \equiv (D_a \,.\, D_\chi) = e^{a\phi}(D_\phi \,.\, D_\chi) = e^{a\phi} d_{\phi\chi}.$$

This is true for all the $m-k$ values of χ. By reciprocation

$$e^{a\phi} = [d^{\phi\chi}]_{m-k}\, d_{a\chi}.$$

This agrees with (IX. 7. 1) and (IX. 7. 2). As $m-k$ will usually be a good deal greater than k, the method is rather of theoretical than of practical interest.

(vi) The elements which we have found to be important in the above processes are the $m-k$ auxiliaries D_ϕ, whose improved values are all 0, and the k elements D^a of the

conjugate set which correspond to the remainder of D_λ. These elements together constitute a set of order m; and we have in fact, in (iii), expressed D_α in terms of this set. As the set is important, it is worth while to see what is its conjugate.

We write
$$E^\lambda \equiv D^\alpha \& D_\phi,$$
where the '&' means that the elements of the two sets of orders k and $m-k$ are combined to form a set of order m. These two partial sets are statistically independent. It follows, by § 3 (ix), that the set conjugate to E^λ is
$$E_\lambda = (d_{\alpha\beta})_k D^\beta \& [d^{\phi\chi}]_{m-k} D_\chi.$$

(vii) But ($d_{\alpha\beta}$ and $d_{\beta\alpha}$ being identical) we have already found that
$$ID_\alpha = (d_{\alpha\beta})_k D^\beta.$$
Hence we get a concise formula for finding ID_α. Let the set conjugate to D_α & D_ϕ be D^α & D^ϕ; and let the set conjugate to D^α & D_ϕ be F_α & F^ϕ. Then $F_\alpha = ID_\alpha$.

(viii) Since ID_α is of the form $D_\alpha - e^{\alpha\phi} D_\phi$, and ID_β is of the form $e_{\beta\gamma} D^\gamma$, and $(D^\gamma \, . \, D_\phi) = 0$, it follows that
$$(ID_\alpha \, . \, ID_\beta) = (D_\alpha \, . \, ID_\beta),$$
and similarly
$$(ID_\alpha \, . \, ID_\beta) = (ID_\alpha \, . \, D_\beta).$$

[NOTE.—This chapter is based on (1) a paper by myself in *Phil. Trans.* (1920), ser. A, vol. 221, pp. 199-237, in which the old notation was used; (2) a paper by Professor Eddington in *Proceedings of the London Mathematical Society*, ser. 2, vol. 20, pp. 213-221, showing how the notation and methods of the tensor calculus can be applied, and making some abbreviations and improvements in the work; and (3) a note by myself, following the above, *ibid.*, pp. 222-224. I have altered the notation a good deal.]

X. TENSORS IN THEORY OF RELATIVITY

X. 1. **Preliminary.**—(i) Tensors are sets* which (1) are functions of a set of co-ordinates ($x_1\, x_2\, x_3 \ldots$) and (2) are subject to certain conditions of transformation when the co-ordinates are transformed.

(ii) In the theory of relativity there are four co-ordinates ($x_1\, x_2\, x_3\, x_4$), so that all the sets are of order 4, and any inner multiplication with regard to a suffix μ involves addition of the products for the values 1, 2, 3, 4 of μ. But this fixing of the number of elements in a set does not affect the general reasoning with regard to the sets, and we can continue to treat them as of order m.

(iii) In VII. 4 we started with a set X^λ, and a set A^λ which is a definite function of X^λ, and we supposed A^λ to be changed as the result of X^λ being changed by linear substitution; and the cases we considered were those in which, throughout all such changes, A^λ varies either directly or reciprocally as X^λ. In VII. 7 we extended the inquiry by taking a set $\mathfrak{A}$ to be a function of two or more single sets, and considered cases in which, when these sets are changed by linear substitution, $\mathfrak{A}$ varies directly or reciprocally as each of the sets. For tensors we have tc consider cases in which the primary substitutions are not necessarily linear. If in place of the original set of co-ordinates x_λ we take a new set x'_λ, which is a function but not necessarily a linear function of x_λ, the ratio which

* It must be remembered that the elements of a set are not necessarily numbers, but may be quantities; and that a set as a whole is something different from its elements. What we usually mean by a tensor is some physical phenomenon represented by a set: but no confusion arises if we call the set itself a tensor.

we have now to consider is not the ratio of x'_ρ to x_λ but the ratio of their differentials, i.e. the partial differential coefficient $\partial x'_\rho/\partial x_\lambda$. When the substitution is linear, this is equal * to x'_ρ/x_λ, so that our treatment of the general case is consistent with our previous treatment of the particular case.

We will begin with the single set, and then go on to sets of higher rank.

(iv) In the case of sets of higher rank, we sometimes have to deal not only with inner products but also with *inner sums*. By the **inner sum**, in such an expression as $A^\rho_{\mu\nu\sigma}$, we mean the result obtained by replacing ρ by σ and summing the values for $\sigma = 1, 2, 3, 4$. It will be seen presently (§ 5 (iv)) that, as the result of the particular notation adopted, the inner products or sums have only to be considered when one of the two letters concerned is an upper suffix and the other is a lower suffix.

X. 2. Single sets (vectors).—(i) Beginning with single sets, we start with a pair which we call x_λ (the set of co-ordinates) and A^λ, or x_a and A^a; A^λ or A^a being a function of x_λ or x_a. (The a here, like the λ, denotes a complete set, not, as in Chapter IX, a partial set.) Connected with these, or arising out of them, is a plurality of pairs x'_λ and A'^λ. But our purview is limited to the cases in which the relation between A^λ and A'^λ is linear, and in which, further, this linear relation is of one of the two following kinds:

(*a*) where

$$\frac{A'^\lambda}{A^a} = \frac{\partial x'_\lambda}{\partial x_a}:$$

* The difficulty mentioned in the note to VIII. 1 (ii) does not arise, because $\partial x'_\rho/\partial x_\lambda$ does not occur absolutely, but (directly or reciprocally) as one of the factors in inner multiplication with regard to λ or ρ.

(*b*) —replacing A^λ and A'^λ by A_λ and A'_λ—where

$$\frac{A'_\lambda}{A_a} = \frac{\partial x_a}{\partial x'_\lambda}.$$

In the cases under (*a*) A^λ is said to be a **contravariant vector**; in the cases under (*b*) A_λ is said to be a **covariant vector.** Here 'vector' is used as meaning a single set which is a tensor.

(ii) It will be seen from VIII. 1 (ii) that, if the relation between x_λ and x'_λ is linear, these become respectively

$$\frac{A'^\lambda}{A^a} = \frac{x'_\lambda}{x_a},$$

$$\frac{A'_\lambda}{A_a} = \frac{x_a}{x'_\lambda},$$

so that in these particular classes of cases A^λ is *contravariant* if A^λ and x_λ are *cogredient*, and A_λ is *covariant* if A_λ and x_λ are *contragredient*.*

X. 3. Other sets.—(i) For sets of higher rank, we are similarly concerned with pairs of partial derivatives

$$\frac{\partial x'_\lambda}{\partial x_a} \text{ and } \frac{\partial x_a}{\partial x'_\lambda},$$

$$\frac{\partial x'_\mu}{\partial x_\beta} \text{ and } \frac{\partial x_\beta}{\partial x'_\mu},$$

etc.; and a set depending on $x_a, x_\beta, \ldots$ is not a tensor unless, when $x_a x_\beta \ldots$ become $x'_\lambda x'_\mu \ldots$, the 'ratio' of the new value of the set to the old value is the product of these partial derivatives, one from each pair. The particular derivatives are indicated by the position of the letters $a\beta\ldots$ or $\mu\nu\ldots$: these are upper suffixes if, so far as the particular variable is concerned, the relation is of the

* It seems desirable to call attention to these classes of cases, as otherwise the tensor terminology may be found rather confusing.

contravariant type, and lower suffixes if the relation is of the covariant type. Thus for double sets (tensors of the second rank) we should have such relations as

$$\frac{A'^{\lambda\mu}}{A^{\alpha\beta}} = \frac{\partial x'_\lambda}{\partial x_\alpha}\frac{\partial x'_\mu}{\partial x_\beta} \text{ (contravariant tensor),}$$

$$\frac{A'_{\lambda\mu}}{A_{\alpha\beta}} = \frac{\partial x_\alpha}{\partial x'_\lambda}\frac{\partial x_\beta}{\partial x'_\mu} \text{ (covariant tensor),}$$

$$\frac{A'^{\mu}_{\lambda}}{A^{\beta}_{\alpha}} = \frac{\partial x_\alpha}{\partial x'_\lambda}\frac{\partial x'_\mu}{\partial x_\beta} \text{ (mixed tensor);}$$

and for a tensor of higher rank we might have such a relation as

$$\frac{A'^{\rho}_{\lambda\mu\nu}}{A^{\delta}_{\alpha\beta\gamma}} = \frac{\partial x_\alpha}{\partial x'_\lambda}\frac{\partial x_\beta}{\partial x'_\mu}\frac{\partial x_\gamma}{\partial x'_\nu}\frac{\partial x'_\rho}{\partial x_\delta}.$$

Two tensors $\mathfrak{A}$ and $\mathfrak{B}$ are said to be of the same **character** if the ratios $\mathfrak{A}'/\mathfrak{A}$ and $\mathfrak{B}'/\mathfrak{B}$ are of the same form.

(ii) For a scalar function of a set or sets the above condition becomes

$$A' = A,$$

so that a scalar (in the general sense) is not a tensor unless it remains constant for all changes in the system of co-ordinates. Such a function as $A_1 + A_2 + A_3 + A_4$, for instance, would not usually be a tensor. For tensor purposes, therefore, 'scalar' practically means 'invariant'.

X. 4. **Reason for limitation.**—The object of limiting the definition of 'tensor' in this way is to ensure that the result of any number of steps, all of the same kind, produced by successive transformations of co-ordinates, shall be the same as if we had passed in one step, also of the same kind, from the initial set of co-ordinates to the final set. That this is in fact ensured is seen from the properties

of the partial derivative of a set. Suppose, for example, that

$$\frac{A'^{\lambda}}{A^{a}} = \frac{\partial x'_{\lambda}}{\partial x_{a}}, \tag{1}$$

and that

$$\frac{A''^{\sigma}}{A'^{\lambda}} = \frac{\partial x''_{\sigma}}{\partial x'_{\lambda}}. \tag{2}$$

Then, by (VII. 2. 2) and (VIII. 3. 1),

$$\frac{A''^{\sigma}}{A^{a}} = \frac{A''^{\sigma}}{A'^{\lambda}}\frac{A'^{\lambda}}{A^{a}} = \frac{\partial x''_{\sigma}}{\partial x'_{\lambda}}\frac{\partial x'_{\lambda}}{\partial x_{a}} = \frac{\partial x''_{\sigma}}{\partial x_{a}}, \tag{3}$$

which is of the same form as (1) and (2).

X. 5. Miscellaneous properties.—The following are some miscellaneous properties which are useful in determining whether a set is a tensor.

(i) The sum (or difference) of two tensors of the same rank and character and the same suffix is a tensor.

[Suppose, for instance, that A^{λ} and B^{λ} are contravariant tensors. Then

$$A'^{\lambda} = \frac{\partial x'_{\lambda}}{\partial x_{a}} A^{a}, \quad B'^{\lambda} = \frac{\partial x'_{\lambda}}{\partial x_{a}} B^{a},$$

and therefore

$$(A'^{\lambda} + B'^{\lambda}) = \frac{\partial x'_{\lambda}}{\partial x_{a}}(A^{a} + B^{a}).]$$

(ii) The product of two tensors is a tensor whose character is the combination of the characters of the two. For example, from

$$A'_{\lambda} = \frac{\partial x_{a}}{\partial x'_{\lambda}} A_{a}, \quad B'^{\mu\nu} = \frac{\partial x'_{\mu}}{\partial x_{\beta}}\frac{\partial x'_{\nu}}{\partial x_{\gamma}} B^{\beta\gamma},$$

we see that

$$(A'_{\lambda}B'^{\mu\nu}) = \frac{\partial x_{a}}{\partial x'_{\lambda}}\frac{\partial x'_{\mu}}{\partial x_{\beta}}\frac{\partial x'_{\nu}}{\partial x_{\gamma}}(A_{a}B^{\beta\gamma}).$$

(iii) An inner product of two tensors, or an inner sum (§ 1 (iv)) of a tensor, taken with regard to suffixes of opposite character, is a tensor.

[Take, for example,

$$A'^{\rho}_{\lambda\mu\nu} = \frac{\partial x_\alpha}{\partial x'_\lambda}\frac{\partial x_\beta}{\partial x_\mu{}'}\frac{\partial x_\gamma}{\partial x'_\nu}\frac{\partial x'_\rho}{\partial x_\delta}A^{\delta}_{\alpha\beta\gamma}.$$

If we replace ρ by ν we have, by (VIII. 3. 3),

$$\frac{\partial x_\gamma}{\partial x'_\nu}\frac{\partial x'_\nu}{\partial x_\delta}A^{\delta}_{\alpha\beta\gamma} = 1^{\gamma}_{\delta}A^{\delta}_{\alpha\beta\gamma} = A^{\gamma}_{\alpha\beta\gamma};$$

and therefore

$$A'^{\nu}_{\lambda\mu\nu} = \frac{\partial x_\alpha}{\partial x'_\lambda}\frac{\partial x_\beta}{\partial x'_\mu}A^{\gamma}_{\alpha\beta\gamma},$$

which satisfies the requirements.]

(iv) If in this last example we had replaced μ by λ, instead of ρ by ν, the expression would have contained

$$\frac{\partial x_\alpha}{\partial x'_\lambda}\frac{\partial x_\beta}{\partial x'_\lambda},$$

which has no general significance.

(v) The derivative of a scalar (§ 3 (ii)) is a covariant vector. Suppose, e.g., that A is a scalar function of x_λ, whose value remains constant (§ 3 (ii)) for all transformations. Then

$$\frac{\partial A'}{\partial x'_\lambda} = \frac{\partial A}{\partial x'_\lambda} = \frac{\partial x_\alpha}{\partial x'_\lambda}\frac{\partial A}{\partial x_\alpha},$$

so that $\partial A/\partial x_\lambda$ is a covariant vector.

(vi) If the inner product of a set $\mathfrak{A}$ by each of $m\ (=4)$ $\left\{\begin{matrix}\text{contravariant}\\ \text{covariant}\end{matrix}\right\}$ vectors is a tensor, then $\mathfrak{A}$ is a tensor, and

is $\left\{\begin{matrix}\text{covariant}\\ \text{contravariant}\end{matrix}\right\}$ as regards x_μ, where μ is the linked suffix.

[Take the case in which the vectors are contravariant, and suppose that $\mathfrak{A} \equiv A_{\mu\nu\ldots}$. Let the m vectors be denoted by $B^{1\mu}, B^{2\mu}, \ldots B^{m\mu}$, or, collectively, by $B^{\lambda\mu}$; the λ not indicating any tensor character. Then, by hypothesis, $B^{\lambda\mu}$ is a tensor as regards x_μ, and $B^{\lambda\mu}A_{\mu\nu\ldots}$ is a tensor as regards $x_\nu \ldots$, so that

$$B^{\lambda\beta} = \frac{\partial x_\beta}{\partial x'_\mu} B'^{\lambda\mu}, \quad B'^{\lambda\mu} A'_{\mu\nu\ldots} = \frac{\partial x_\gamma}{\partial x'_\nu} \ldots B^{\lambda\beta} A_{\beta\gamma\ldots}.$$

From these we deduce

$$B'^{\lambda\mu} A'_{\mu\nu\ldots} = B'^{\lambda\mu} \frac{\partial x_\beta}{\partial x'_\mu} \frac{\partial x_\gamma}{\partial x'_\nu} \ldots A_{\beta\gamma\ldots},$$

whence, by division by $B'^{\lambda\mu}$ (see VI. 9 (v)),

$$A'_{\mu\nu\ldots} = \frac{\partial x_\beta}{\partial x'_\mu} \frac{\partial x_\gamma}{\partial x'_\nu} \ldots A_{\beta\gamma\ldots}.$$

Hence $A_{\mu\nu\ldots}$ is a tensor and is covariant as regards x_μ. The case in which the vectors are covariant can be dealt with in the same way.]

APPENDIX

PRODUCT OF DETERMINANTS

In IV. 5 we have taken as the standard form for product of two determinants—the order being now reduced from 3 to 2, for economy of printing—

$$\begin{vmatrix} a_1 b_1 \\ a_2 b_2 \end{vmatrix} \times \begin{vmatrix} \alpha_1 \beta_1 \\ \alpha_2 \beta_2 \end{vmatrix} = \begin{vmatrix} a_1\alpha_1 + a_2\beta_1 & b_1\alpha_1 + b_2\beta_1 \\ a_1\alpha_2 + a_2\beta_2 & b_1\alpha_2 + b_2\beta_2 \end{vmatrix}$$

In this form, the element in the qth column and rth row of the result is the 'inner product' of the qth column of the first determinant and the rth row of the second. By interchanges of columns and rows, and also by changing the order of multiplication, we get seven other forms, all constructed according to this rule. The eight forms can be set out as follows:

$$\begin{vmatrix} a_1 b_1 \\ a_2 b_2 \end{vmatrix} \times \begin{vmatrix} \alpha_1 \beta_1 \\ \alpha_2 \beta_2 \end{vmatrix} = \begin{vmatrix} a_1\alpha_1 + a_2\beta_1 & b_1\alpha_1 + b_2\beta_1 \\ a_1\alpha_2 + a_2\beta_2 & b_1\alpha_2 + b_2\beta_2 \end{vmatrix} \qquad (1)$$

$$\begin{vmatrix} a_1 b_1 \\ a_2 b_2 \end{vmatrix} \times \begin{vmatrix} \alpha_1 \alpha_2 \\ \beta_1 \beta_2 \end{vmatrix} = \begin{vmatrix} a_1\alpha_1 + a_2\alpha_2 & b_1\alpha_1 + b_2\alpha_2 \\ a_1\beta_1 + a_2\beta_2 & b_1\beta_1 + b_2\beta_2 \end{vmatrix} \qquad (2)$$

$$\begin{vmatrix} a_1 a_2 \\ b_1 b_2 \end{vmatrix} \times \begin{vmatrix} \alpha_1 \beta_1 \\ \alpha_2 \beta_2 \end{vmatrix} = \begin{vmatrix} a_1\alpha_1 + b_1\beta_1 & a_2\alpha_1 + b_2\beta_1 \\ a_1\alpha_2 + b_1\beta_2 & a_2\alpha_2 + b_2\beta_2 \end{vmatrix} \qquad (3)$$

$$\begin{vmatrix} a_1 a_2 \\ b_1 b_2 \end{vmatrix} \times \begin{vmatrix} \alpha_1 \alpha_2 \\ \beta_1 \beta_2 \end{vmatrix} = \begin{vmatrix} a_1\alpha_1 + b_1\alpha_2 & a_2\alpha_1 + b_2\alpha_2 \\ a_1\beta_1 + b_1\beta_2 & a_2\beta_1 + b_2\beta_2 \end{vmatrix} \qquad (4)$$

$$\begin{vmatrix} \alpha_1 \beta_1 \\ \alpha_2 \beta_2 \end{vmatrix} \times \begin{vmatrix} a_1 b_1 \\ a_2 b_2 \end{vmatrix} = \begin{vmatrix} a_1\alpha_1 + b_1\alpha_2 & a_1\beta_1 + b_1\beta_2 \\ a_2\alpha_1 + b_2\alpha_2 & a_2\beta_1 + b_2\beta_2 \end{vmatrix} \qquad (5)$$

$$\begin{vmatrix} \alpha_1 \alpha_2 \\ \beta_1 \beta_2 \end{vmatrix} \times \begin{vmatrix} a_1 b_1 \\ a_2 b_2 \end{vmatrix} = \begin{vmatrix} a_1\alpha_1 + b_1\beta_1 & a_1\alpha_2 + b_1\beta_2 \\ a_2\alpha_1 + b_2\beta_1 & a_2\alpha_2 + b_2\beta_2 \end{vmatrix} \qquad (6)$$

$$\begin{vmatrix} \alpha_1 \beta_1 \\ \alpha_2 \beta_2 \end{vmatrix} \times \begin{vmatrix} a_1 a_2 \\ b_1 b_2 \end{vmatrix} = \begin{vmatrix} a_1\alpha_1 + a_2\alpha_2 & a_1\beta_1 + a_2\beta_2 \\ b_1\alpha_1 + b_2\alpha_2 & b_1\beta_1 + b_2\beta_2 \end{vmatrix} \qquad (7)$$

$$\begin{vmatrix} \alpha_1 \alpha_2 \\ \beta_1 \beta_2 \end{vmatrix} \times \begin{vmatrix} a_1 a_2 \\ b_1 b_2 \end{vmatrix} = \begin{vmatrix} a_1\alpha_1 + a_2\beta_1 & a_1\alpha_2 + a_2\beta_2 \\ b_1\alpha_1 + b_2\beta_1 & b_1\alpha_2 + b_2\beta_2 \end{vmatrix} \qquad (8)$$

It will be seen that the last four are the transposed of the first four, but in the reverse order; i. e. the transposed of (1) (2) (3) (4) are (8) (7) (6) (5).

In the double-suffix notation these become, the order of (5)–(8) being reversed:

$$|d_{qr}| \times |e_{qr}| = |d_{q\lambda}e_{\lambda r}| \quad (1) \qquad |e_{rq}| \times |d_{rq}| = |d_{r\lambda}e_{\lambda q}| \quad (8)$$
$$|d_{qr}| \times |e_{rq}| = |d_{q\lambda}e_{r\lambda}| \quad (2) \qquad |e_{qr}| \times |d_{rq}| = |d_{r\lambda}e_{q\lambda}| \quad (7)$$
$$|d_{rq}| \times |e_{qr}| = |d_{\lambda q}e_{\lambda r}| \quad (3) \qquad |e_{rq}| \times |d_{qr}| = |d_{\lambda r}e_{\lambda q}| \quad (6)$$
$$|d_{rq}| \times |e_{rq}| = |d_{\lambda q}e_{r\lambda}| \quad (4) \qquad |e_{qr}| \times |d_{qr}| = |d_{\lambda r}e_{q\lambda}| \quad (5)$$

It must be borne in mind that in each of these statements q refers to the column and r to the row; e. g. (6) means that

$$\begin{vmatrix} e_{11} e_{12} \\ e_{21} e_{22} \end{vmatrix} \times \begin{vmatrix} d_{11} d_{21} \\ d_{12} d_{22} \end{vmatrix} = \begin{vmatrix} d_{\lambda 1}e_{\lambda 1} & d_{\lambda 1}e_{\lambda 2} \\ d_{\lambda 2}e_{\lambda 1} & d_{\lambda 2}e_{\lambda 2} \end{vmatrix}.$$

INDEX OF SYMBOLS

GENERAL INDEX

Under any heading, separate entries are usually in the order of occurrence, not in alphabetical order.

AN INTRODUCTION TO COMBINATORY ANALYSIS

BY

MAJOR P. A. MACMAHON,
D.Sc., Sc.D., LL.D., F.R.S.
Member of St John's College, Cambridge

PREFACE

THIS little book is intended to be an Introduction to the two volumes of *Combinatory Analysis* which were published by the Cambridge University Press in 1915–16. It has appeared to me to be necessary from the circumstance that some of my mathematical critics have found that the presentation of the general problem through the medium of the algebra of symmetric functions is difficult or troublesome reading. I was reminded that the great Euler wrote a famous algebra which was addressed to his man-servant, and had the object of anticipating and removing every conceivable difficulty and obscurity. Posterity gives the verdict that, in accomplishing this he was wonderfully successful.

From a general point of view it seems to me there is advantage on the one hand in explaining a complicated if not difficult matter to an untrained mind, and on the other in propounding a simple theory for the benefit of those who are highly trained. In this way certain vantage points may be reached which are not commonly attainable by the usual plan of addressing students in a style which is in proportion to their attainments. The advantage which has been spoken of accrues both to the writer and to the reader. The writer for example is likely to be led to points of view of whose existence he was previously unaware or aware of only sub-consciously. In attempting what is here proposed it is inevitable that much must be written that will appear to the reader to be self-evident and unworthy of statement. The intention is by a succession of such statements to arrive at facts which, by a quicker progression, would be difficult or troublesome to grasp. It is in analogy with a succession of likenesses of a person taken at small intervals of time such that little or no difference can be detected between any two successive pictures but between pictures taken at

considerable intervals there is but a mere resemblance. The subject-matter of the book shews I believe that the algebra of symmetric functions and an important part of Combinatory Analysis are beautifully adapted to one another, and if I have succeeded in making that clear to the reader I shall be satisfied that the object of the book has been attained.

My grateful thanks are due to Professor J. E. A. Steggall, M.A. for much helpful criticism during the composition of the book.

P. A. M.

February, 1920.

TABLE OF CONTENTS

CHAPTER I

ELEMENTARY THEORY OF SYMMETRIC FUNCTIONS

CHAPTER II

OPENING OF THE THEORY OF DISTRIBUTIONS

CHAPTER III

DISTRIBUTION INTO DIFFERENT BOXES

CHAPTER IV

DISTRIBUTION WHEN OBJECTS AND BOXES ARE EQUAL IN NUMBER

CHAPTER V

DISTRIBUTIONS OF GIVEN SPECIFICATION

CHAPTER VI

THE MOST GENERAL CASE OF DISTRIBUTION

CHAPTER I

ELEMENTARY THEORY OF SYMMETRIC FUNCTIONS

1. A great part of Combinatory Analysis may be based upon the algebra of Symmetric Functions, and it is therefore necessary to have some clear definitions and simple properties of such functions before us.

An algebraic function of a number of numerical magnitudes is said to be Symmetrical if it be unaltered when *any* two of the magnitudes are interchanged. In algebra such magnitudes (or quantities) are denoted by letters of the alphabet.

Restricting ourselves to those functions which are rational it is clear, for example, that the simple sum of the quantities $\alpha, \beta, \gamma, \ldots \nu$, n in number, is such a function. For the sum

$$\alpha + \beta + \gamma + \ldots + \nu$$

is unaltered when any selected pair of the letters is interchanged. For this symmetric function, of which α is the type, we adopt the short-hand

$$\Sigma\alpha.$$

Again, another symmetric function is

$$\alpha^i + \beta^i + \gamma^i + \ldots + \nu^i,$$

because the enunciated conditions of symmetry are just as clearly satisfied as in the particular case $i = 1$.

We may denote this function by

$$\Sigma\alpha^i,$$

the representative or typical term being alone put in evidence. This last expression includes all the integral symmetric functions, the representative term of which involves one only of the quantities. If we are not restricted to integral functions the representative term may be any rational function of α. For example

$$\Sigma \frac{\alpha^s}{1 - a\alpha^i} = \frac{\alpha^s}{1 - a\alpha^i} + \frac{\beta^s}{1 - a\beta^i} + \frac{\gamma^s}{1 - a\gamma^i} + \ldots + \frac{\nu^s}{1 - a\nu^i},$$

but we are, in most cases, concerned with the symmetric functions which are integral as well as rational.

The function $\Sigma\alpha^i$ is the sum of the ith powers of the quantities. It takes a leading part in the algebra of the functions.

The laws of this algebra do not depend upon the absolute magnitudes of the quantities $\alpha, \beta, \gamma, \ldots \nu$, so that usually it is not necessary to specify these quantities. Various notations have been adopted with the object of eliminating the actual magnitudes from consideration. Thus $\Sigma\alpha^i$ is sometimes denoted by s_i; meaning thereby the sum of the ith powers of magnitudes which it is not needful to specify either in magnitude or (very often) in number. Others realising that in the algebra they have to deal entirely with the number i have denoted the same function by

$$(i),$$

viz. the number i in round brackets. This notation is of the greater importance because, as will become evident, it can be extended readily to rational and integral functions in general. Not only so; it is fundamentally important because it supplies the connecting link between the algebra of symmetric functions and theories which deal with numbers only and not with algebraic quantities.

2. Proceeding to functions whose representative terms involve two quantities, the simplest we find to be

$$\alpha\beta + \alpha\gamma + \beta\gamma + \ldots + \mu\nu,$$

which involves each of the $\frac{1}{2}n(n-1)$ combinations, two together, of the n quantities. It is visibly symmetrical.

This is denoted in conformity with the conventional notation by

$$\Sigma\alpha\beta,$$

or by

$$(11),$$

the function being completely given when n is known.

Every function is considered to have a *weight*, which is equal to the sum of the numbers that, in the last notation, appear in the brackets.

Thus the functions (i), (11) have the weights i, 2 respectively.

When a number is repeated in brackets it is convenient to use repetitional exponents. Thus

(11) is frequently written in the form (1^2).

Of the weight one we have the single function

$$(1);$$

of the weight two, the two functions

$$(2),\ (1^2).$$

Observe that two functions present themselves because two objects can either be taken in one lot comprising both objects, or in two lots, one object in each lot. We express this by saying that the number 2 has two partitions. We have thus, of the weight two, a function corresponding to each partition of 2.

3. In the notation of the Theory of the Partition of Numbers the partitions of the number 2 are denoted by (2), (1^2). It is for this reason that the notation we are employing for symmetric functions is termed 'The Partition Notation.' Similarly in correspondence with the three partitions of 3, viz. (3), (21), (1^3), we have the symmetric functions

$$\Sigma\alpha^3,\ \Sigma\alpha^2\beta,\ \Sigma\alpha\beta\gamma$$

of the weight 3.

Of symmetric functions whose representative terms involve two of the n quantities we have the two types in which the repetitional exponents are alike, or different,

$$\Sigma\alpha^i\beta^i \equiv \alpha^i\beta^i + \alpha^i\gamma^i + \beta^i\gamma^i + \ldots + \mu^i\nu^i = (i^2),$$

$$\Sigma\alpha^i\beta^j \equiv \alpha^i\beta^j + \alpha^j\beta^i + \ldots + \mu^i\nu^j + \mu^j\nu^i = (ij),$$

involving $\frac{1}{2}n(n-1)$ and $n(n-1)$ terms respectively.

It is now an easy step to the function

$$\Sigma\alpha_1^{i_1}\alpha_2^{i_2}\alpha_3^{i_3}\ldots\alpha_s^{i_s},$$

wherein we have replaced the quantities $\alpha, \beta, \gamma, \ldots \nu$ by the suffixed series $\alpha_1, \alpha_2, \alpha_3, \ldots \alpha_s$.

In the partition notation we write the function

$$(i_1 i_2 i_3 \ldots i_s),$$

where of course s cannot be greater than n.

It involves a number of terms which can be computed when we know the equalities that occur between the numbers $i_1, i_2, i_3, \ldots i_s$.

If we are thinking only of numbers, $(i_1 i_2 i_3 \ldots i_s)$ is a partition of a number $N = i_1 + i_2 + i_3 + \ldots + i_s$, and since a partition of N is defined to be any collection of positive integers whose sum is N we may consider the numbers $i_1, i_2, i_3, \ldots i_s$ to be in descending order of magnitude. These numbers are called the Parts of the partition and the partition is said to have s parts.

The series of functions denoted by (i) for different integer values of i constitute a first important set. They are sometimes called one-part functions.

4. A second important set is constituted by those functions which are denoted by partitions in which only unity appears as a part. It is

$$(1),\ (1^2),\ (1^3),\ \ldots\ (1^n),$$

or

$$\Sigma\alpha_1,\ \Sigma\alpha_1\alpha_2,\ \Sigma\alpha_1\alpha_2\alpha_3,\ \ldots\ \Sigma\alpha_1\alpha_2\alpha_3\ldots\alpha_n.$$

These are sometimes called unitary functions.

The set is particularly connected with the Theory of Algebraic Equations because

$$(x-\alpha)(x-\beta)(x-\gamma)\ldots(x-\nu)$$
$$= x^n - \Sigma\alpha\,.\,x^{n-1} + \Sigma\alpha\beta\,.\,x^{n-2} - \Sigma\alpha\beta\gamma\,.\,x^{n-3} + \ldots,$$

the last term being $\pm\Sigma\alpha\beta\gamma\ldots\nu$, according as n is even or uneven. Hence considering the equation

$$x^n - a_1x^{n-1} + a_2x^{n-2} - a_3x^{n-3} + \ldots + (-)^n a_n = 0,$$

and supposing the n roots to be

$$\alpha,\ \beta,\ \gamma,\ \ldots\ \nu,$$

it is clear that

$$x^n - a_1x^{n-1} + a_2x^{n-2} - \ldots + (-)^n a_n$$
$$= x^n - \Sigma\alpha\,.\,x^{n-1} + \Sigma\alpha\beta\,.\,x^{n-2} - \ldots + (-)^n\alpha\beta\gamma\ldots\nu,$$

and we at once deduce the relations

$$a_1 = \Sigma\alpha,$$
$$a_2 = \Sigma\alpha\beta,$$
$$a_3 = \Sigma\alpha\beta\gamma,$$
$$\ldots\ldots\ldots\ldots$$
$$a_n = \alpha\beta\gamma\ldots\nu.$$

These functions are frequently called 'elementary' symmetric functions because they arise in this simple manner.

It is sometimes convenient, undoubtedly, to regard the quantities $\alpha, \beta, \gamma, \ldots \nu$ as being the roots of an equation, the left-hand side of which involves the elementary functions with alternately positive and negative signs, but the notion is not essential to the study of the subject of symmetric functions.

5. There is a third important series of functions.

Of the weight w there are functions which in the partition notation are denoted by partitions of the number w.

There is one function corresponding to every such partition.

Such a function, since it is denoted by a single partition, is called a Monomial Symmetric Function.

If we add all such functions which have the same weight we obtain, algebraically speaking, all the products w together of the quantities $\alpha, \beta, \gamma, \dots \nu$, repetitions permissible.

Such a sum is called the Homogeneous Product-Sum of weight w of the n quantities.

It is usually denoted by h_w.

We have

$$h_1 = (1) = \Sigma\alpha,$$
$$h_2 = (2) + (1^2) = \Sigma\alpha^2 + \Sigma\alpha\beta,$$
$$h_3 = (3) + (21) + (1^3) = \Sigma\alpha^3 + \Sigma\alpha^2\beta + \Sigma\alpha\beta\gamma,$$

and so forth.

We have before us the three sets of functions

$$s_1, s_2, s_3, \dots s_\nu, \dots,$$
$$a_1, a_2, a_3, \dots a_\nu,$$
$$h_1, h_2, h_3, \dots h_\nu, \dots.$$

The first and third sets contain an infinite number of members, but the second set only involves n members where n is the number of the quantities $\alpha, \beta, \gamma, \dots$.

6. The identity of Art. 4 which connects the functions $a_1, a_2, a_3, \dots$ with $\alpha, \beta, \gamma, \dots$ may be written, by putting $\frac{1}{y}$ for x,

$$1 - a_1 y + a_2 y^2 - \dots + (-)^n a_n y^n \equiv (1 - \alpha y)(1 - \beta y) \dots (1 - \nu y),$$

or in the form

$$\frac{1}{1 - a_1 y + a_2 y^2 - \dots + (-)^n a_n y^n} \equiv \frac{1}{(1 - \alpha y)(1 - \beta y) \dots (1 - \nu y)}.$$

If we expand the last fraction in ascending powers of y, we obtain, in the first place,

$$\begin{aligned}&1\\ &+ (\alpha + \beta + \gamma + \dots + \nu)\, y\\ &+ (\alpha^2 + \beta^2 + \gamma^2 + \dots + \nu^2 + \alpha\beta + \alpha\gamma + \beta\gamma + \dots + \mu\nu)\, y^2\\ &+ (\alpha^3 + \beta^3 + \gamma^3 + \dots + \nu^3 + \alpha^2\beta + \alpha\beta^2 + \dots + \mu^2\nu + \mu\nu^2 + \alpha\beta\gamma + \alpha\beta\delta + \dots + \lambda\mu\nu)\, y^3\\ &+ \dots.\end{aligned}$$

It is clear that the coefficient of y^w is the homogeneous product-sum of weight w, so that we may write

$$\frac{1}{1 - a_1 y + a_2 y^2 - \dots + (-)^n a_n y^n} \equiv 1 + h_1 y + h_2 y^2 + \dots + h_w y^w + \dots,$$

an identity.

Thence we obtain

$$\{1 - a_1 y + a_2 y^2 - \ldots + (-)^n a_n y^n\}(1 + h_1 y + h_2 y^2 + \ldots + h_w y^w + \ldots) = 1.$$

Since this is an identity we may multiply out the left-hand side and equate the coefficients of the successive powers of y to zero; obtaining

$$\begin{aligned}
&h_1 - a_1 = 0,\\
&h_2 - a_1 h_1 + a_2 = 0,\\
&h_3 - a_1 h_2 + a_2 h_1 - a_3 = 0,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&h_n - a_1 h_{n-1} + a_2 h_{n-2} - \ldots + (-)^n a_n = 0,\\
&h_{n+1} - a_1 h_n + a_2 h_{n-1} - \ldots + (-)^n a_n h_1 = 0,\\
&h_{n+2} - a_1 h_{n+1} + a_2 h_n - \ldots + (-)^n a_n h_2 = 0,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

relations which enable us to express any function h_w in terms of members of the series $a_1, a_2, a_3, \ldots a_n$.

7. In the applications to combinatory analysis it usually happens that we may regard n as being indefinitely great and then the relations are simply

$$\begin{aligned}
&h_1 - a_1 = 0,\\
&h_2 - a_1 h_1 + a_2 = 0,\\
&h_3 - a_1 h_2 + a_2 h_1 - a_3 = 0,\\
&\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

continued indefinitely.

The before-written identity now becomes

$$(1 - a_1 y + a_2 y^2 - a_3 y^3 + \ldots \textit{ad inf.})(1 + h_1 y + h_2 y^2 + h_3 y^3 + \ldots \textit{ad inf.}) \equiv 1,$$

and herein writing $-y$ for y and transposing the factors we find

$$(1 - h_1 y + h_2 y^2 - h_3 y^3 + \ldots \textit{ad inf.})(1 + a_1 y + a_2 y^2 + a_3 y^3 + \ldots \textit{ad inf.}) \equiv 1,$$

an identity which is derivable from the former by interchange of the symbols a and h.

There is thus perfect symmetry between the symbols and it follows as a matter of course that in any relation connecting the quantities $a_1, a_2, a_3, \ldots$ with the quantities $h_1, h_2, h_3, \ldots$ we are at liberty to interchange the symbols a, h. This interesting fact can be at once verified in the case of the relations $h_1 - a_1 = 0$, etc.

Solving these equations we find

$$\begin{aligned}
&h_1 = a_1,\\
&h_2 = a_1^2 - a_2,\\
&h_3 = a_1^3 - 2a_1 a_2 + a_3,
\end{aligned}$$

and as shewn in works upon algebra

$$h_n = \Sigma\,(-)^{n+\pi_1+\pi_2+\ldots+\pi_k}\frac{(\pi_1+\pi_2+\ldots+\pi_k)!}{\pi_1!\,\pi_2!\ldots\pi_k!}\,a_1^{\pi_1}a_2^{\pi_2}\ldots a_k^{\pi_k},$$

where $\pi_1!$ denotes the factorial of π_1 and

$$\pi_1 + 2\pi_2 + 3\pi_3 + \ldots + k\pi_k = n,$$

the summation being taken for all sets of positive integers $\pi_1, \pi_2, \ldots \pi_k$ which satisfy this equation.

By interchange of symbols we pass to the relations

$$a_1 = h_1,$$
$$a_2 = h_1^2 - h_2,$$
$$a_3 = h_1^3 - 2h_1h_2 + h_3,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_n = \Sigma\,(-)^{n+\pi_1+\pi_2+\ldots+\pi_k}\frac{(\pi_1+\pi_2+\ldots+\pi_k)!}{\pi_1!\,\pi_2!\ldots\pi_k!}\,h_1^{\pi_1}h_2^{\pi_2}\ldots h_k^{\pi_k}.$$

8. It is shewn in works upon algebra that the relations between the symbols $s_1, s_2, s_3, \ldots$ and the symbols $a_1, a_2, a_3, \ldots$ are

$$s_1 = a_1,$$
$$s_2 = a_1^2 - 2a_2,$$
$$s_3 = a_1^3 - 3a_1a_2 + 3a_3,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$s_n = \Sigma\,(-)^{n+\pi_1+\pi_2+\ldots+\pi_k}\frac{(\pi_1+\pi_2+\ldots+\pi_k-1)!\,n}{\pi_1!\,\pi_2!\ldots\pi_k!}\,a_1^{\pi_1}a_2^{\pi_2}\ldots a_k^{\pi_k}.$$
$$a_1 = s_1,$$
$$2!\,a_2 = s_1^2 - s_2,$$
$$3!\,a_3 = s_1^3 - 3s_1s_2 + 2s_3,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$n!\,a_n = \Sigma(-)^{n+\pi_1+\pi_2+\ldots+\pi_k}\frac{n!}{1^{\pi_1}.\,2^{\pi_2}\ldots k^{\pi_k}.\,\pi_1!\,\pi_2!\ldots\pi_k!}\,s_1^{\pi_1}s_2^{\pi_2}\ldots s_k^{\pi_k};$$

also between the symbols $s_1, s_2, s_3, \ldots$ and $h_1, h_2, h_3, \ldots$

$$s_1 = h_1,$$
$$s_2 = -(h_1^2 - 2h_2),$$
$$s_3 = h_1^3 - 3h_1h_2 + 3h_3,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$s_n = \Sigma\,(-)^{\pi_1+\pi_2+\ldots+\pi_k+1}\frac{(\pi_1+\pi_2+\ldots+\pi_k-1)!\,n}{\pi_1!\,\pi_2!\ldots\pi_k!}\,h_1^{\pi_1}h_2^{\pi_2}\ldots h_k^{\pi_k}.$$

$$h_1 = s_1,$$
$$2!\,h_2 = s_1^2 + s_2,$$
$$3!\,h_3 = s_1^3 + 3s_1s_2 + 2s_3,$$
$$\dots\dots\dots\dots\dots\dots$$
$$n!\,h_n = \Sigma \frac{n!}{1^{\pi_1}.2^{\pi_2}\dots k^{\pi_k}.\pi_1!\,\pi_2!\dots\pi_k!} s_1^{\pi_1}s_2^{\pi_2}\dots s_k^{\pi_k}.$$

These are the principal properties of symmetric functions that will be of use.

9. If we take any assemblage of letters such as $\alpha\alpha\beta\gamma$ and are not concerned with the order in which these letters are written, we have a 'Combination' of the letters. If however the order in which the letters are written be taken into account, we have a 'Permutation' of the letters. In the present case we have twelve permutations, viz.

$\alpha\alpha\beta\gamma$	$\alpha\alpha\gamma\beta$	$\alpha\beta\alpha\gamma$	$\alpha\beta\gamma\alpha$	$\alpha\gamma\alpha\beta$	$\alpha\gamma\beta\alpha$
$\beta\alpha\alpha\gamma$	$\beta\alpha\gamma\alpha$	$\beta\gamma\alpha\alpha$	$\gamma\alpha\alpha\beta$	$\gamma\alpha\beta\alpha$	$\gamma\beta\alpha\alpha$

10. In a similar manner if we take any collection of integers which add up to a given integer we have as above defined (Art. 3) a partition of the given number; here no account is taken of the order in which the parts of the partition may be written; but if order has to be taken into account each way of writing the parts is called a 'Composition' of the number, such composition appertaining to the particular partition which is involved. Thus of the number 9, $3321 \equiv 3^221$ is a partition which gives rise to the twelve compositions:

3321	3312	3231	3213	3132	3123
2331	2313	2133	1332	1323	1233

and it will be noticed that the compositions which appertain to the partition 3321 of the number 9 are in correspondence with the permutations of the combination $\alpha\alpha\beta\gamma$.

Moreover, in general the compositions which appertain to any given partition of a number are in correspondence with the permutations of a certain combination of letters.

11. In pursuing the main object of this book, namely the study of the algebra of symmetric functions together with those theories of combination, permutation, arrangement, order and distribution which are summed up in the title 'Combinatory Analysis,' it is important to have some specific rules for arranging the order in which the terms of algebraic expressions are written down.

A monomial symmetric function, as defined in Art. 5, is the sum of a number of different combinations of the same type. In writing out at length these combinations of quantities $\alpha, \beta, \ldots \nu$ we may adopt what is called the 'dictionary' (or alternatively 'alphabetical') order.

In any particular combination we write the α's first, then the β's, and so forth; also one combination is given priority of another if, considering the two combinations to be words, the dictionary would give the one word before the other.

This allusion to the dictionary, with which all persons are familiar, seems to define shortly and clearly the principle of order usually adopted. Thus we write the monomial function of four quantities $\alpha, \beta, \gamma, \delta$

$$\Sigma\alpha^2\beta\gamma$$

in the order
$$\alpha^2\beta\gamma + \alpha^2\beta\delta + \alpha^2\gamma\delta + \alpha\beta^2\gamma + \alpha\beta^2\delta + \alpha\beta\gamma^2 + \alpha\beta\delta^2 + \alpha\gamma^2\delta + \alpha\gamma\delta^2 + \beta^2\gamma\delta + \beta\gamma^2\delta + \beta\gamma\delta^2,$$

the dictionary order being in evidence both in each combination and in the order of the combinations.

Another order is sometimes useful. We may have, in each combination, the repetitional numbers always in the same order as they appear in the representative combination but *subject to this rule*, the letters in dictionary order.

The combinations thus written would then be arranged in dictionary order. Thus we might write

$$\begin{aligned}&\Sigma\alpha^2\beta\gamma\\ &= \alpha^2\beta\gamma + \alpha^2\beta\delta + \alpha^2\gamma\delta + \beta^2\alpha\gamma + \beta^2\alpha\delta + \beta^2\gamma\delta\\ &\quad + \gamma^2\alpha\beta + \gamma^2\alpha\delta + \gamma^2\beta\delta + \delta^2\alpha\beta + \delta^2\alpha\gamma + \delta^2\beta\gamma.\end{aligned}$$

12. Again, frequently we have to write out at length the permutations of a given combination of letters. Here again it is usual to adopt the dictionary order, each permutation being regarded as a dictionary word. Thus the twelve permutations of $\alpha\alpha\beta\gamma$ are written

$$\begin{array}{cccccc}\alpha\alpha\beta\gamma & \alpha\alpha\gamma\beta & \alpha\beta\alpha\gamma & \alpha\beta\gamma\alpha & \alpha\gamma\alpha\beta & \alpha\gamma\beta\alpha\\ \beta\alpha\alpha\gamma & \beta\alpha\gamma\alpha & \beta\gamma\alpha\alpha & \gamma\alpha\alpha\beta & \gamma\alpha\beta\alpha & \gamma\beta\alpha\alpha.\end{array}$$

13. When we have to write out symmetric functions, of the same weight, expressed in partition notation, we usually adopt numerical order, the meaning of which will be clear from the example

$$\begin{aligned}&h_6\\ &= (6) + (51) + (42) + (411) + (33) + (321) + (3111)\\ &\qquad + (222) + (2211) + (21111) + (111111),\end{aligned}$$

where in each term the largest number available is written first, the next largest second, and so on; and in ordering the partitions numbers in descending order of magnitude are in the same relation as are the successive letters of the alphabet in dictionary order. The alternative method is to adopt the numerical order subject to the rule that the partitions are to be arranged in ascending order in regard to the *number of parts* involved.

This would give the expression

$$\begin{aligned} &h_6 \\ &= (6) + (51) + (42) + (33) + (411) + (321) + (222) \\ &\quad + (3111) + (2211) + (21111) + (111111). \end{aligned}$$

When we have to write out the compositions associated with a given partition we adopt numerical order.

Thus associated with the partition (321) of the number 6 we have the compositions

$$(321),\ (312),\ (231),\ (213),\ (132),\ (123).$$

CHAPTER II

OPENING OF THE THEORY OF DISTRIBUTIONS

14. From the principles set forth in the concluding articles of Chapter I we can realise a definite way of expressing the result of algebraical multiplication.

Suppose that we have to form the product of a number of algebraical expressions each of which involves (say) three terms. The expressions are supposed to be given in a definite order from left to right. This order will be determined, usually, by the circumstances.

Let the factors be n in number, and, in the given definite order, denoted by

$$(a_1 + b_1 + c_1)(a_2 + b_2 + c_2)(a_3 + b_3 + c_3) \ldots (a_{n-1} + b_{n-1} + c_{n-1})(a_n + b_n + c_n),$$

where three terms are involved in each factor merely for the sake of simplicity.

To obtain a term of the product we select a term (any term) from each factor and place them in contact in the order in which they have been selected; the factors being dealt with in order from left to right. The term of the product, thus reached, may involve one, two or three of the symbols a, b, c according to the way that the selection is carried out. To place the terms (3^n in number) thus arrived at in a definite order we make our selections in such wise that the terms produced are in dictionary order. Thus the first three terms will be

$$a_1 a_2 a_3 \ldots a_{n-1} a_n,$$
$$a_1 a_2 a_3 \ldots a_{n-1} b_n,$$
$$a_1 a_2 a_3 \ldots a_{n-1} c_n,$$

and the last three

$$c_1 c_2 c_3 \ldots c_{n-1} a_n,$$
$$c_1 c_2 c_3 \ldots c_{n-1} b_n,$$
$$c_1 c_2 c_3 \ldots c_{n-1} c_n.$$

15. As an example, consider the development of $(\alpha + \beta)^n$.

Writing down the n factors

$$(\alpha + \beta)(\alpha + \beta)(\alpha + \beta) \ldots (\alpha + \beta)(\alpha + \beta),$$

the multiplication, according to rule, gives for a few terms

$$\alpha^n + \alpha^{n-1}\beta + \alpha^{n-2}\beta\alpha + \alpha^{n-2}\beta\beta + \alpha^{n-3}\beta\alpha\alpha$$
$$+ \alpha^{n-3}\beta\alpha\beta + \alpha^{n-3}\beta\beta\alpha + \alpha^{n-3}\beta\beta\beta + \ldots,$$

and the complete product for $n = 4$ is

$$\alpha^4 + \alpha^3\beta + \alpha^2\beta\alpha + \alpha^2\beta^2 + \alpha\beta\alpha^2 + \alpha\beta\alpha\beta + \alpha\beta^2\alpha + \alpha\beta^3$$
$$+ \beta\alpha^3 + \beta\alpha^2\beta + \beta\alpha\beta\alpha + \beta\alpha\beta\beta + \beta\beta\alpha\alpha + \beta\beta\alpha\beta + \beta\beta\beta\alpha + \beta^4.$$

The combinations which appear are

$$\alpha^4,\ \alpha^3\beta,\ \alpha^2\beta^2,\ \alpha\beta^3,\ \beta^4,$$

and the product, as obtained by rule, involves all the permutations of these combinations, and no other terms.

The terms of the product are visibly in dictionary order and from the way in which the multiplication has been carried out each of the combinations necessarily appears as many times as it possesses permutations; so that when the terms are assembled so as to yield the formula of the binomial theorem

$$(\alpha + \beta)^4 = \alpha^4 + 4\alpha^3\beta + 6\alpha^2\beta^2 + 4\alpha\beta^3 + \beta^4,$$

each numerical coefficient denotes the number of permutations of the combination of letters to which it is attached. The same remark can be made in regard to the general formula

$$(\alpha + \beta)^n = \alpha^n + \binom{n}{1}\alpha^{n-1}\beta + \binom{n}{2}\alpha^{n-2}\beta^2 + \ldots + \binom{n}{1}\alpha\beta^{n-1} + \beta^n.$$

16. We proceed to another order of ideas by connecting the theory above sketched with a Distribution into different Boxes.

Suppose that we have four different (that is distinguishable) boxes A_1, A_2, A_3, A_4 arranged in order from left to right

A_1	A_2	A_3	A_4
α	α	β	β
α	β	α	β
α	β	β	α
β	β	α	β
β	α	β	α
β	α	α	α

and let us consider the selections of factor terms that were made in forming the product combination $\alpha^2\beta^2$.

For the first selection we took the terms $\alpha, \alpha, \beta, \beta$ from the first, second, third and fourth factors respectively. Place these letters in the four boxes A_1, A_2, A_3, A_4 respectively. Proceed in the same way for

each of the six selections that produce the combination $\alpha^2\beta^2$. We obtain the successive lines of letters shewn in the above scheme. We observe that we have distributed the four letters $\alpha, \alpha, \beta, \beta$ into the four different boxes, one letter into each box, in every possible way, and that reading the lines of letters from left to right one such distribution corresponds to each permutation of the combination. From the mode of term selection, to form the product, each permutation must occur and there can be no other distributions into the boxes except those which correspond to the permutations.

We thus see that, if the binomial expression

$$(\alpha + \beta)^n$$

be expanded, the coefficient of the term $\alpha^m\beta^{n-m}$ is equal

(i) to the number of permutations of the combination $\alpha^m\beta^{n-m}$,

(ii) to the number of ways in which the letters of the combination $\alpha^m\beta^{n-m}$ can be distributed into n different boxes, one letter into each box.

17. By precisely the same argument we reach the conclusion that if the multinomial expression

$$(\alpha_1 + \alpha_2 + \alpha_3 + \dots + \alpha_n)^i$$

be expanded, the coefficient of the term

$$\alpha_1^{i_1}\alpha_2^{i_2}\alpha_3^{i_3}\dots\alpha_s^{i_s}$$

is equal

(i) to the number of permutations of the combination $\alpha_1^{i_1}\alpha_2^{i_2}\dots\alpha_s^{i_s}$,

(ii) to the number of ways in which the letters of the combination $\alpha_1^{i_1}\alpha_2^{i_2}\dots\alpha_s^{i_s}$ can be distributed into i different boxes, so that each box contains one letter.

We now remark that since $\alpha_1 + \alpha_2 + \alpha_3 + \dots + \alpha_n$ is a symmetric function of the n quantities, the expression

$$(\alpha_1 + \alpha_2 + \alpha_3 + \dots + \alpha_n)^i$$

is also a symmetric function. Hence every term of the expansion which is also a term of the function

$$\Sigma\alpha_1^{i_1}\alpha_2^{i_2}\alpha_3^{i_3}\dots\alpha_s^{i_s} \equiv (i_1 i_2 i_3 \dots i_s)$$

must appear with the same coefficient.

Hence we may say that when

$$(\alpha_1 + \alpha_2 + \alpha_3 + \dots + \alpha_n)^i$$

is expanded the coefficient of symmetric function

$$\Sigma\alpha_1^{i_1}\alpha_2^{i_2}\dots\alpha_s^{i_s}$$

is equal

(i) to the number of permutations of the combination $\alpha_1^{i_1}\alpha_2^{i_2}\ldots\alpha_s^{i_s}$,

(ii) to the number of ways in which the letters of the combination $\alpha_1^{i_1}\alpha_2^{i_2}\ldots\alpha_s^{i_s}$ can be distributed into i different boxes, one letter into each box.

18. We have spoken of the distribution of letters into boxes. The letters may represent objects or things and it is often more convenient to speak of the distribution of objects rather than of letters. The objects are sufficiently specified by the repetitional numbers $i_1, i_2, \ldots i_s$. We can therefore properly define the objects distributed or permuted by saying that they have a Specification

$$(i_1 i_2 \ldots i_s)$$

which is necessarily some partition of i, the whole number of objects.

Also the number of ways in which the distribution into boxes can take place depends upon the identities that may exist between them. The boxes being such that there are no identities—in fact representable by

$$A_1, A_2, \ldots A_i,$$

we have only to regard the repetitional numbers, which in this case consist of i units. The Box Specification is therefore

$$(1^i),$$

and the distribution which we have had under view may be described as of objects specified by $(i_1 i_2 i_3 \ldots i_s)$ into boxes specified by (1^i), where

$$i_1 + i_2 + i_3 + \ldots + i_s = i.$$

19. The actual number of permutations or distributions is readily obtained. If the objects be all different, or in other words of specification (1^i), we may select an object for box A_1 in i ways; from the $i-1$ remaining objects we can select our object for box A_2 in $i-1$ ways; consequently we can use the boxes A_1, A_2 in $i(i-1)$ ways; similarly we can use the boxes A_1, A_2, A_3 in $i(i-1)(i-2)$ ways, and the whole of the boxes in $i(i-1)(i-2)\ldots 2.1$ or in $i!$ ways. Hence there are $i!$ ways of distributing objects of specification (1^i) into boxes of specification (1^i). Now suppose i_1 of the objects are identical. In any distribution certain boxes i_1 in number will contain the same objects and if we now replace these similar objects by different objects we find that we can do this in $i_1!$ ways corresponding to i_1 permutations of the i_1 different objects. Hence the former distributions are $i_1!$ times as numerous as the latter, and therefore we find that the latter distributions can take place in

$$\frac{i!}{i_1!} \text{ ways.}$$

Similarly if other i_2 objects be identical the number of distributions is

$$\frac{i!}{i_1!\,i_2!},$$

and finally if the specification of the objects be

$$(i_1 i_2 \ldots i_s)$$

the number of distributions into boxes of specification (1^i) is

$$\frac{i!}{i_1!\,i_2!\ldots i_s!}.$$

This number therefore enumerates the permutations of objects of specification $(i_1 i_2 \ldots i_s)$.

20. In the partition notation the multinomial theorem may be written

$$(1)^i = \Sigma \frac{i!}{i_1!\,i_2!\ldots i_s!}(i_1 i_2 \ldots i_s),$$

the summation being for every partition of the number i.

It will be observed that the multinomial theorem involves the enumeration of the permutations of all combinations of letters that it is possible to form. For this reason it is often said to be the 'Generating Function' for the enumeration of permutations. Since it also enumerates certain distributions it may be said to be the 'Distribution Function' for the distribution of objects into boxes of specification (1^i), one object being placed in each box.

21. From this first very elementary case of distribution we can at once derive an interesting corollary.

Suppose that we have to distribute i objects into $i+j$ different boxes, so that the box specification is (1^{i+j}) subject to the condition that no box is to contain more than one object. It is clear that in any distribution there must be j empty boxes, and that we may place in each of them one of j new and identical objects.

These j new objects have the specification (j). Hence the problem before us is transformed into that of distributing objects of specification

$$(i_1 i_2 \ldots i_s j)$$

into boxes of specification (1^{i+j}).

The objects and boxes being now equi-numerous we have the case already considered and can see that the number of distributions is

$$\frac{(i+j)!}{i_1!\,i_2!\ldots i_s!\,j!}.$$

22. The reader will now observe that we can also pass from the latter to the former distribution, and that just as we can add any number to the specification of the objects in order to equalise the objects and boxes; so conversely, if we are given any specification of i objects and boxes of specification (1^i) we can cancel *any part* from the specification of the objects without altering the number of the distributions. Thus the distributions of objects $(i_1 i_2 \ldots i_s)$ into boxes (1^i) are equi-numerous with the distributions of objects

$$(i_1 i_2 \ldots i_{r-1} i_{r+1} \ldots i_s)$$

into boxes (1^i). Here the number i_r is cancelled from the object specification, and r may be any of the numbers $1, 2, \ldots s$.

As a simple example we find that objects of the specifications (21^2), (1^2), (21) have equi-numerous distributions into four different boxes, not more than one object in each box. These are

A_1	A_2	A_3	A_4
α	α	β	γ
α	α	γ	β
α	β	α	γ
α	β	γ	α
α	γ	α	β
α	γ	β	α

A_1	A_2	A_3	A_4
β	α	α	γ
β	α	γ	α
β	γ	α	α
γ	α	α	β
γ	α	β	α
γ	β	α	α

A_1	A_2	A_3	A_4
α	β		
α		β	
α			β
	α	β	
	α		β
		α	β

A_1	A_2	A_3	A_4
β	α		
β		α	
β			α
	β	α	
	β		α
		β	α

A_1	A_2	A_3	A_4
α	α	β	
α	α		β
α	β	α	
α		α	β
α	β		α
α		β	α

A_1	A_2	A_3	A_4
β	α	α	
	α	α	β
β	α		α
	α	β	α
β		α	α
	β	α	α

23. In the case of the binomial theorem another interpretation may be given to the coefficients.

Writing

$$(\alpha+\beta)^n = \binom{n}{0}\alpha^n + \binom{n}{1}\alpha^{n-1}\beta + \dots + \binom{n}{m}\alpha^{n-m}\beta^m + \dots + \binom{n}{n}\beta^n,$$

it has been shewn that the coefficient $\binom{n}{m}$ enumerates the permutations of the combination $\alpha^{n-m}\beta^m$.

We can shew that the same number enumerates the number of ways of selecting m letters from an assemblage of n different letters. For consider the combination $\alpha^3\beta^2$ and its ten permutations

$$\begin{array}{ccccc} \alpha\alpha\alpha\beta\beta & \alpha\alpha\beta\alpha\beta & \alpha\alpha\beta\beta\alpha & \alpha\beta\alpha\alpha\beta & \alpha\beta\alpha\beta\alpha \\ \alpha\beta\beta\alpha\alpha & \beta\alpha\alpha\alpha\beta & \beta\alpha\alpha\beta\alpha & \beta\alpha\beta\alpha\alpha & \beta\beta\alpha\alpha\alpha \end{array}$$

When an α is in the sth place from the left of the permutation substitute for it the suffixed α, α_s; thus obtaining, omitting the letters β entirely, the ten combinations

$$\begin{array}{ccccc} \alpha_1\alpha_2\alpha_3 & \alpha_1\alpha_2\alpha_4 & \alpha_1\alpha_2\alpha_5 & \alpha_1\alpha_3\alpha_4 & \alpha_1\alpha_3\alpha_5 \\ \alpha_1\alpha_4\alpha_5 & \alpha_2\alpha_3\alpha_4 & \alpha_2\alpha_3\alpha_5 & \alpha_2\alpha_4\alpha_5 & \alpha_3\alpha_4\alpha_5 \end{array}$$

which constitute the ten combinations three together that can be formed from the letters of the combination $\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5$. If we had operated similarly with the letters β we would have reached the ten combinations two together that can be formed from the same combination of five letters, viz.

$$\begin{array}{ccccc} \alpha_4\alpha_5 & \alpha_3\alpha_5 & \alpha_3\alpha_4 & \alpha_2\alpha_5 & \alpha_2\alpha_4 \\ \alpha_2\alpha_3 & \alpha_1\alpha_5 & \alpha_1\alpha_4 & \alpha_1\alpha_3 & \alpha_1\alpha_2 \end{array}$$

and we have no difficulty in realising that the number $\binom{n}{m} \equiv \binom{n}{n-m}$ enumerates the combinations m or $n-m$ together that can be formed from n different letters.

24. So far we have been concerned mainly with a distribution into different boxes, one object only being placed in each box. The results have been trivial, but they have supplied a connecting link between combinatory analysis and the algebra of symmetric functions. It will be shewn in what follows that this relationship can be greatly extended to the mutual advantage of combinatory analysis and the symmetrical algebra.

We proceed in the first place by removing the restriction that each box is to contain only one object. We consider distributions in which

the number of boxes is less than the number of objects, so that some box or boxes must contain more than one object. We may join the issue in two ways. We may precisely define the distribution and then seek its connexion with the algebra; or we may set forth some combination of symmetric functions, which we can see will lead to a distribution of the required kind, and then seek to define the corresponding distribution. For the present we adopt the latter procedure and inquire into the development of the function

$$(1^2)^i \equiv (\Sigma\alpha\beta)^i,$$

which for four quantities may be written

$$(\alpha\beta+\alpha\gamma+\alpha\delta+\beta\gamma+\beta\delta+\gamma\delta)^i.$$

We carry out the multiplication of the i factors according to the process explained in Art. 14. The complete product is clearly a symmetric function, expressible as a linear function of monomials, of the weight $2i$, of the form

$$\Sigma\alpha_1^{i_1}\alpha_2^{i_2}\ldots\alpha_s^{i_s} \equiv (i_1 i_2 \ldots i_s).$$

The monomial function just written will appear with a certain coefficient. What is the meaning of that coefficient in the theory of distributions?

In the process of multiplication we take any combination of two letters from the first factor with any combination from the second, and so on, until finally we take any combination from the ith factor and, assembling the letters thus obtained, we obtain, we suppose, a combination of letters

$$\alpha_1^{i_1}\alpha_2^{i_2}\alpha_3^{i_3}\ldots\alpha_s^{i_s}.$$

Associated with this step in the multiplication we take i different boxes, that is to say of specification (1^i),

$$A_1 A_2 A_3 \ldots A_i$$

corresponding to the i factors of the product in order from left to right and place the two-letter combinations which have been selected from the 1st, 2nd, ... ith factors in the boxes $A_1, A_2, \ldots A_i$ respectively. We thus arrive at a distribution of the letters in $\alpha_1^{i_1}\alpha_2^{i_2}\ldots\alpha_s^{i_s}$ into i different boxes, subject to the sole condition that each box is to contain two different letters; or, as we may say, letters of specification (1^2). Making a similar distribution in correspondence with every selective step in the multiplication, which produces the combination $\alpha_1^{i_1}\alpha_2^{i_2}\ldots\alpha_s^{i_s}$, we reach a set of distributions which constitute the whole number of ways of distributing a definite set of objects of specification $(i_1 i_2 \ldots i_s)$ into boxes of specification (1^i) in such wise that every box

contains objects of specification (1^2). Since all the terms included in $\Sigma\alpha_1^{i_1}\alpha_2^{i_2}\ldots\alpha_s^{i_s}$, regarded as denoting objects, have the same specification we say that the number of the distributions above defined is equal to the coefficient of symmetric function

$$(i_1 i_2 \ldots i_s)$$

in the development of the symmetric function

$$(1^2)^i.$$

Some examples supply simple verifications.

By ordinary multiplication we find that

$$(\alpha\beta + \alpha\gamma + \alpha\delta + \beta\gamma + \beta\delta + \gamma\delta)^2 = \Sigma\alpha^2\beta^2 + 2\Sigma\alpha^2\beta\gamma + 6\alpha\beta\gamma\delta,$$

or

$$(1^2)^2 = (2^2) + 2\,(21^2) + 6\,(1^4).$$

The distributions which give the three coefficients 1, 2, 6 are

A_1	A_2
$\alpha\beta$	$\alpha\beta$

A_1	A_2
$\alpha\beta$	$\alpha\gamma$
$\alpha\gamma$	$\alpha\beta$

A_1	A_2
$\alpha\beta$	$\gamma\delta$
$\alpha\gamma$	$\beta\delta$
$\alpha\delta$	$\beta\gamma$
$\beta\gamma$	$\alpha\delta$
$\beta\delta$	$\alpha\gamma$
$\gamma\delta$	$\alpha\beta$

Again, by developing

$$(\Sigma\alpha\beta)^3 \equiv (1^2)^3$$

we find a term

$$15\Sigma\alpha^2\beta^2\gamma\delta \equiv 15\,(2^2 1^2).$$

The fifteen distributions are

A_1	A_2	A_3
$\alpha\beta$	$\alpha\gamma$	$\beta\delta$
$\alpha\beta$	$\beta\delta$	$\alpha\gamma$
$\alpha\gamma$	$\alpha\beta$	$\beta\delta$
$\alpha\gamma$	$\beta\delta$	$\alpha\beta$
$\beta\delta$	$\alpha\beta$	$\alpha\gamma$
$\beta\delta$	$\alpha\gamma$	$\alpha\beta$

A_1	A_2	A_3
$\alpha\beta$	$\alpha\delta$	$\beta\gamma$
$\alpha\beta$	$\beta\gamma$	$\alpha\delta$
$\alpha\delta$	$\alpha\beta$	$\beta\gamma$
$\alpha\delta$	$\beta\gamma$	$\alpha\beta$
$\beta\gamma$	$\alpha\beta$	$\alpha\delta$
$\beta\gamma$	$\alpha\delta$	$\alpha\beta$

A_1	A_2	A_3
$\alpha\beta$	$\alpha\beta$	$\gamma\delta$
$\alpha\beta$	$\gamma\delta$	$\alpha\beta$
$\gamma\delta$	$\alpha\beta$	$\alpha\beta$

There is no simple expression for the general coefficient in the development of $(\Sigma\alpha\beta)^i$; but when i is not too large there is a method of arriving at the value of any desired coefficient which will be given at a later stage.

25. We pass on to consider the symmetric function product

$$(\Sigma\alpha\beta)^i(\Sigma\alpha)^j \equiv (1^2)^i(1)^j.$$

We write out the i factors followed by the j factors and proceed to obtain one term in the development by taking one combination of two letters from each of the first i factors and one letter from each of the last j factors. Assembling the letters so obtained we reach, suppose, a combination of $2i+j$ letters

$$\alpha_1^{p_1}\alpha_2^{p_2}\dots\alpha_s^{p_s}.$$

In correspondence with the selective process that has resulted in this combination we take $i+j$ different boxes, so that the box specification is (1^{i+j})

$$A_1A_2A_3\dots A_iB_1B_2B_3\dots B_j.$$

We place the two-letter combinations that were selected from the 1st, 2nd, ... ith factors in the boxes $A_1, A_2, \dots A_i$ respectively; and the single letters that were selected from the last j factors in the boxes $B_1, B_2, \dots B_j$ respectively. If we make a similar distribution for every case in which the selective process in the multiplication results in the combination $\alpha_1^{p_1}\alpha_2^{p_2}\dots\alpha_s^{p_s}$ we will have obtained every distribution of a definite set of objects of specification $(p_1p_2\dots p_s)$ with boxes of specification (1^{i+j}) in such wise that in regard to i of the boxes $A_1, A_2, \dots A_i$ each box contains objects of specification (1^2), and in the remaining boxes $B_1, B_2, \dots B_j$ each box contains a single object. Hence we gather that distributions so specified are enumerated by the coefficient of the function $(p_1p_2\dots p_s)$ in the development of the product

$$(1^2)^i(1)^j.$$

In the distribution above defined the reader must notice that objects of specifications (1^2), (1) are restricted to the boxes $A_1, A_2, \dots$; $B_1, B_2, \dots$ respectively. This implies that the boxes being in a definite order the $i+j$ combinations of objects are only allowed $i!j!$ permutations; that is to say that no exchange of combinations of objects of different specifications is allowed to take place. *If* such exchange be permitted $(i+j)!$ permutations between the combinations of objects may take place. The function that by its development enumerates the distributions must now be multiplied by

$$(i+j)! \div i!j! \equiv \binom{i+j}{i},$$

and we have the theorem:—

"If objects of specification $(p_1p_2\dots p_s)$ be distributed into boxes of specification (1^{i+j}) in such wise that i of the boxes (unspecified) receive

objects of specification (1^2) and the remaining boxes objects of specification (1), the number of distributions is equal to the coefficient of the function $(p_1 p_2 \ldots p_s)$ in the development of the function

$$\binom{i+j}{i}(1^2)^i(1)^j."$$

As an example it is found that

$$\binom{4}{2}(1^2)^2(1)^2 = \ldots + 48\,(321) + \ldots.$$

The 48 distributions are

	A_1	A_2	A_3	A_4
the 12 permutations of	$\alpha\beta$	$\alpha\beta$	α	γ
24 "	$\alpha\beta$	$\alpha\gamma$	α	β
12 "	$\alpha\beta$	$\beta\gamma$	α	α

26. It is quite evident that the process by which we have reached this connecting link between distributions and the expansion of symmetric function products is of general application. The selective process is in correspondence with distribution when the factors of the symmetric function products are *any monomial symmetric functions whatever.*

For consider the product

$$(\Sigma\alpha_1^{m_1}\alpha_2^{m_2}\ldots\alpha_t^{m_t})^i\,(\Sigma\alpha_1^{n_1}\alpha_2^{n_2}\ldots\alpha_u^{n_u})^j \equiv (m_1 m_2 \ldots m_t)^i (n_1 n_2 \ldots n_u)^j.$$

We write out the i factors followed by the j factors and obtain one term in the development by taking one term from each of the $i+j$ factors. The i terms from the first i factors are each of them combinations of specification $(m_1 m_2 \ldots m_t)$. The j terms from the last j factors are each of them of specification $(n_1 n_2 \ldots n_u)$. The assemblage of $i+j$ terms is, suppose,

$$\alpha_1^{p_1}\alpha_2^{p_2}\ldots\alpha_s^{p_s} \text{ of specification } (p_1 p_2 \ldots p_s).$$

In correspondence with the selective process we take $i+j$ boxes of specification (1^{i+j})

$$A_1 A_2 \ldots A_i \quad B_1 B_2 \ldots B_j.$$

We place the combinations that have been selected from the first i factors in the boxes A respectively and the remaining combinations in the boxes B.

If we make a similar distribution for every case in which the selective process results in the combination $\alpha_1^{p_1}\alpha_2^{p_2}\ldots\alpha_s^{p_s}$ we will have obtained every distribution of a definite set of objects of specification $(p_1 p_2 \ldots p_s)$ into boxes of specification (1^{i+j}) subject to the condition

that the combinations of specifications $(m_1 m_2 \dots m_t)$, $(n_1 n_2 \dots n_u)$ must be placed in the boxes A, B respectively.

Removing this condition we find as before a theorem :—

"If objects of specification $(p_1 p_2 \dots p_s)$ be distributed into boxes of specification (1^{i+j}) in such wise that i of the boxes (unspecified) receive objects of specification $(m_1 m_2 \dots m_t)$ and the remaining boxes objects of specification $(n_1 n_2 \dots n_u)$, the number of distributions is equal to the coefficient of the function $(p_1 p_2 \dots p_s)$ in the development of the function

$$\binom{i+j}{i}(m_1 m_2 \dots m_t)^i (n_1 n_2 \dots n_u)^j."$$

27. The same reasoning applies when any number of monomial symmetric functions are multiplied together and we may enunciate the general theorem :—

"If objects of specification $(p_1 p_2 \dots p_s)$ be distributed into boxes of specification $(1^{i+j+k+\dots})$ in such wise that i unspecified boxes receive objects of specification $(m_1 m_2 \dots m_t)$, j other unspecified boxes objects of specification $(n_1 n_2 \dots n_u)$, k other unspecified boxes objects of specification $(o_1 o_2 \dots o_v)$, etc., the number of distributions is equal to the coefficient of the function $(p_1 p_2 \dots p_s)$ in the development of the function

$$\frac{(i+j+k+\dots)!}{i!\,j!\,k!\dots}(m_1 m_2 \dots m_t)^i (n_1 n_2 \dots n_u)^j (o_1 o_2 \dots o_v)^k \dots"$$

Verifications may be made by means of the formula

$$(\Sigma\alpha^2\beta\gamma)(\Sigma\alpha\beta\gamma) = \Sigma\alpha^3\beta^2\gamma^2 + 2\Sigma\alpha^3\beta^2\gamma\delta + 6\Sigma\alpha^3\beta\gamma\delta\epsilon$$
$$+ 3\Sigma\alpha^2\beta^2\gamma^2\delta + 6\Sigma\alpha^2\beta^2\gamma\delta\epsilon + 10\Sigma\alpha^2\beta\gamma\delta\epsilon\theta,$$

otherwise written

$$(21^2)(1^3) = (32^2) + 2(321^2) + 6(31^4) + 3(2^3 1) + 6(2^2 1^3) + 10(21^5).$$

28. As it is important to be able to obtain readily the numerical values of such coefficients, we will subject this particular development to examination with the object of deducing general laws in the algebra of symmetric functions.

Suppose that the symmetric functions appertain to an unlimited number of quantities $\alpha, \beta, \gamma, \dots$ and expand each side of the identity in powers of one of them, say α. The function (32^2) or $\Sigma\alpha^3\beta^2\gamma^2$ involves some terms which do not contain α; terms such as $\beta^3\gamma^2\delta^2$ for example. The aggregate of these terms is (32^2) regarded as appertaining to the set of quantities $\beta, \gamma, \delta, \dots$, the original set with the omission of α. The function involves no terms containing the first power of α, but it has terms such as $\alpha^2\beta^3\gamma^2$ which contain the second power of α, the aggregate

of which is $\alpha^2(32)$, if (32) now appertains to the set $\beta, \gamma, \delta, \ldots$. Lastly it involves terms $\alpha^3(2^2)$, where (2^2) refers to the set $\beta, \gamma, \delta, \ldots$.

Hence we may write

$$(32^2) = (32^2)' + \alpha^2(32)' + \alpha^3(2^2)',$$

the dashed round bracket denoting that the symmetric functions refer to the deficient set of quantities $\beta, \gamma, \delta, \ldots$.

Similarly

$$\begin{aligned}
(321^2) &= (321^2)' + \alpha(321)' + \alpha^2(31^2)' + \alpha^3(21^2)', \\
(31^4) &= (31^4)' + \alpha(31^3)' + \alpha^3(1^4)', \\
(2^31) &= (2^31)' + \alpha(2^3)' + \alpha^2(2^21)', \\
(2^21^3) &= (2^21^3)' + \alpha(2^21^2)' + \alpha^2(21^3)', \\
(21^5) &= (21^5)' + \alpha(21^4)' + \alpha^2(1^5)'.
\end{aligned}$$

The right-hand side of the identity may therefore be written

$$\begin{aligned}
(32^2)' &+ 2(321^2)' + 6(31^4)' + 3(2^31)' + 6(2^21^3)' + 10(21^5)' \\
&+ \alpha\{2(321)' + 6(31^3)' + 3(2^3)' + 6(2^21^2)' + 10(21^4)'\} \\
&+ \alpha^2\{(32)' + 2(31^2)' + 3(2^21)' + 6(21^3)' + 10(1^5)'\} \\
&+ \alpha^3\{(2^2)' + 2(21^2)' + 6(1^4)'\}.
\end{aligned}$$

As regards the left-hand side, since

$$\begin{aligned}
(21^2) &= (21^2)' + \alpha(21)' + \alpha^2(1^2)', \\
(1^3) &= (1^3)' + \alpha(1^2)',
\end{aligned}$$

we find that

$$\begin{aligned}
(21^2)(1^3) = (21^2)'(1^3)' &+ \alpha\{(21)'(1^3)' + (21^2)'(1^2)'\} \\
&+ \alpha^2\{(1^2)'(1^3)' + (21)'(1^2)'\} + \alpha^3(1^2)'(1^2)'.
\end{aligned}$$

Now equating the coefficients of like powers of α (omitting the case α^0) and suppressing the dashes to the round brackets by converting the set of quantities $\beta, \gamma, \delta, \ldots$ into the set $\alpha, \beta, \gamma, \ldots$ through writing $\alpha, \beta, \gamma, \ldots$ for $\beta, \gamma, \delta, \ldots$ respectively, we obtain the derived formulae

$$\begin{aligned}
(21)(1^3) + (21^2)(1^2) &= 2(321) + 6(31^3) + 3(2^3) + 6(2^21^2) + 10(21^4), \\
(1^2)(1^3) + (21)(1^2) &= (32) + 2(31^2) + 3(2^21) + 6(21^3) + 10(1^5), \\
(1^2)^2 &= (2^2) + 2(21^2) + 6(1^4).
\end{aligned}$$

Thus we can derive, from any given identity, a number of other identities of lower weights. The very simple process is that of expansion in ascending powers of the quantity α.

We observe that the coefficient of α^m in any monomial function is obtained by merely deleting the part m from the partition which denotes the function; if the part m be not present the coefficient is zero. Observe also that in the product $(21^2)(1^3)$ the highest power of α that presents itself is 3 because 2, 1 are the largest parts in the factors respectively

and $2+1=3$. It follows at once that the coefficient of α^3 in the product is found by simply obliterating the first or largest part in each factor. We thus arrive at the coefficient $(1^2)^2$. Thus from the original identity $(21^2)(1^3)=(32^2)+2(321^2)+6(31^4)+$ other terms which involve no part, in the partitions, as large as 3, we derive, at sight, the new identity

$$(1^2)^2=(2^2)+2(21^2)+6(1^4).$$

From this we discover immediately new theorems in distribution. As an example, since

$$2(21^2)(1^3)=\ldots+12(31^4)+\ldots,$$
$$(1^2)^2=\ldots+6(1^4)+\ldots,$$

we can assert that the number of distributions of objects of specification (31^4) into boxes of specification (1^2) in such wise that the boxes contain objects of specification (21^2) and (1^3) is twice the number of distributions of objects of specification (1^4) into boxes of specification (1^2) in such wise that both boxes contain objects of specification (1^2).

Examination of the distributions verifies this conclusion and the theory we are now discussing might have been entirely based upon a study of the distributions.

29. In order to facilitate the process of taking the coefficients of α^m in a symmetric function it is convenient to adopt a mathematical shorthand. Let the symbol

$$D_m,$$

placed before any symmetric function, stand for the phrase

'the coefficients of α^m in.'

Then when D_m is prefixed to a monomial function expressed in the partition notation, the result is the deletion of the part m from the partition; if the part m be not present the result is zero; if m itself be zero the result is to leave the function unaltered or, as we may say, to multiply the function by unity. For example

$$D_3(32^2)=(2^2);\ D_2(32^2)=(32);\ D_3(3)=1;$$
$$D_4(32^2)=D_1(32^2)=0;\ D_0(32^2)=(32^2)^*.$$

* It should be stated that the reader who is acquainted with the differential calculus will realise that D_m is effectively a partial differential operator of the order m which is expressible by means of symmetric functions in a variety of ways and, in particular, in terms of the elementary functions (1), (1^2), (1^3), ... which have been denoted above also by $a_1, a_2, a_3, \ldots$.

It was brought to light in 1883, *Proc. Lond. Math. Soc.*, by James Hammond and is freely used in 'Combinatory Analysis' and in many researches by the author which have been published in Scientific Journals during the past thirty years. The methods of the calculus are not necessary for this elementary exposition and the requisite properties of the symbol will be set forth without its aid.

30. Any symmetric function F may be written in ascending powers of the quantity α in the form

$$D_0 F + \alpha D_1 F + \alpha^2 D_2 F + \dots,$$

in accordance with the definition of $D_m F$.

Hence the product of two functions F_1, F_2 is

$$(D_0 F_1 + \alpha D_1 F_1 + \alpha^2 D_2 F_1 + \dots)(D_0 F_2 + \alpha D_1 F_2 + \alpha^2 D_2 F_2 + \dots),$$

or

$$\begin{aligned} & D_0 F_1 D_0 F_2 \\ & + \alpha\, (D_0 F_1 D_1 F_2 + D_1 F_1 D_0 F_2) \\ & + \alpha^2 (D_0 F_1 D_2 F_2 + D_1 F_1 D_1 F_2 + D_2 F_1 D_0 F_2) \\ & + \alpha^3 (D_0 F_1 D_3 F_2 + D_1 F_1 D_2 F_2 + D_2 F_1 D_1 F_2 + D_3 F_1 D_0 F_2) \\ & + \dots. \end{aligned}$$

Moreover

$$F_1 F_2 = D_0 (F_1 F_2) + \alpha D_1 (F_1 F_2) + \alpha^2 D_2 (F_1 F_2) + \dots.$$

Whence comparing the coefficients of α, α^2, α^3, ...,

$$\begin{aligned} D_1 (F_1 F_2) &= D_0 F_1 \,.\, D_1 F_2 + D_1 F_1 \,.\, D_0 F_2, \\ D_2 (F_1 F_2) &= D_0 F_1 \,.\, D_2 F_2 + D_1 F_1 \,.\, D_1 F_2 + D_2 F_1 \,.\, D_0 F_2, \\ D_3 (F_1 F_2) &= D_0 F_1 \,.\, D_3 F_2 + D_1 F_1 \,.\, D_2 F_2 + D_2 F_1 \,.\, D_1 F_2 + D_3 F_1 \,.\, D_0 F_2, \\ & \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ D_m (F_1 F_2) &= \sum_{s=0}^{s=m} D_s F_1 \,.\, D_{m-s} F_2, \end{aligned}$$

where on the right-hand side there is a term in correspondence with every composition (see Art. 10) of the number m, zero counting as a part. There are visibly $m + 1$ terms, but usually fewer than $m + 1$ will materialise because by the rules of operation many terms may vanish.

Similarly if we require the coefficients of α^m in the product of three functions

$$F_1 F_2 F_3,$$

the performance of the symbol D_m will involve a term

$$D_s F_1 \,.\, D_t F_2 \,.\, D_{m-s-t} F_3,$$

because one step in the multiplication is to find the coefficients of α^s, α^t, α^{m-s-t} in F_1, F_2, F_3 respectively, and then to multiply the three coefficients together.

Hence

$$D_m (F_1 F_2 F_3) = \sum_{s=0} \sum_{t=0} D_s F_1 \,.\, D_t F_2 \,.\, D_{m-s-t} F_3.$$

Since s, t, $m - s - t$ is a composition of the number into three parts, zero counting as a part, the symbol D_m breaks up into as many triads

of symbols as the number m possesses compositions into three parts, zero counting as a part. The reader will have little difficulty in proving that the number of these compositions is

$$D_m(1+a+a^2+\ldots)^3 = D_m(1-a)^{-3} = \binom{m+2}{2} = \binom{m+2}{m}.$$

In general, when the symbol D_m is prefixed to a product of i symmetric functious, it breaks up into as many i-ads of symbols as the number i possesses compositions into i parts, zero counting as a part. The number of such compositions is

$$D_m(1-a)^{-i} = \binom{m+i-1}{i-1} = \binom{m+i-1}{m}.$$

31. We can now see the importance of the study of the symbol, for evidently we can repeatedly operate with it, varying its suffix as may be desired, until a positive integer or zero is reached, and thus solve the problem of the multiplication of symmetric functions upon which the present view of combinatory analysis depends. For consider the product

$$(21^2)(1^3),$$

we have

$$D_3(21^2)(1^3) = D_2(21^2) \,.\, D_1(1^3) = (1^2)(1^2),$$
$$D_3D_2(21^2)(1^3) = D_2(1^2)(1^2) = D_1(1^2) \,.\, D_1(1^2) = (1)(1),$$
$$D_3D_2D_1(21^2)(1^3) = D_0(1) \,.\, D_1(1) + D_1(1) \,.\, D_0(1) = 2(1),$$

and finally

$$D_3D_2D_1^2(21^2)(1^3) = 2D_1(1) = 2.$$

Now we may write

$$(21^2)(1^3) = \ldots + C(321^2) + \ldots,$$

so that operating upon both sides with $D_3D_2D_1^2$ the right-hand side becomes C since every other term is reduced to zero by the operation. The calculation above shews that the left-hand side becomes 2 by the operation. Hence $C = 2$, and

$$(21^2)(1^3) = \ldots + 2\,(321^2) + \ldots.$$

We can in this way calculate the result of the product of any number of monomial functions and thus evaluate the number which enumerates a well-defined distribution of objects into boxes.

CHAPTER III

DISTRIBUTION INTO DIFFERENT BOXES

32. The theory set forth in the foregoing chapters enables us to make a great advance in combinatory analysis.

We are now able to attack the following problem.

Objects of any given specification are to be distributed into m different boxes, i.e. of specification (1^m); in how many ways can the distribution be made?

First consider the case of two boxes, denoted by A_1, A_2, and let the objects be w in number. It has been shewn in Art. 26 that if the specification of the objects be $(p_1 p_2 \dots p_s)$ and the boxes are obliged to contain objects of specifications $(m_1 m_2 \dots m_t)$, $(n_1 n_2 \dots n_u)$, both specifications appearing, one in each box, the enumerating symmetric function product is $2(m_1 m_2 \dots m_t)(n_1 n_2 \dots n_u)$ or $(m_1 m_2 \dots m_t)^2$ if the partitions $(m_1 m_2 \dots m_t)$, $(n_1 n_2 \dots n_u)$ be identical.

We have merely to develop the product and seek the coefficient of the function $(p_1 p_2 \dots p_s)$.

We now abolish the restriction and substitute another, viz. that the boxes are to receive, one of them w_1 objects and the other w_2 objects. We have

$$w_1 + w_2 = w;$$

the w objects may have any specification and the w_1 and w_2 objects may have any specifications consistent with the condition that the assemblage of w_1 and w_2 objects must have the same specification as the w objects. If the specification of the w objects which are to be distributed be unknown the w_1 and w_2 objects may have specifications denoted by any partitions of w_1 and w_2 respectively. The w_1 objects may have therefore all specifications included in the function h_{w_1}, the w_2 objects all those included in the function h_{w_2}. If we form the functions

$$2h_{w_1}h_{w_2} \text{ or } h_{w_1}^2,$$

according as w_1, w_2 are not or are equal, we obtain, upon multiplication, terms of the forms

$$2(m_1 m_2 \dots m_t)(n_1 n_2 \dots n_u) \text{ or } (m_1 m_2 \dots m_t)^2,$$

and it has been shewn already that these functions enumerate, on development, the distributions which are associated with the particular specifications $(m_1 m_2 \dots m_t)$, $(n_1 n_2 \dots n_u)$.

As an example let us distribute 4 objects into two different boxes so that one box, unspecified, contains 3 objects and the other box 1 object.

We have

$$\begin{aligned} 2h_3h_1 &= 2\{(3)+(21)+(1^3)\}(1) \\ &= 2(4)+4(31)+4(2^2)+6(21^2)+8(1^4), \end{aligned}$$

leading to the conclusion that objects of specification (21^2) can be distributed in 6 ways and similarly when the objects have other specifications.

The distributions for all of the cases are:

Spec.	(4)		(31)		(2^2)		(21^2)		(1^4)	
	A_1	A_2	A_1	A_2	A_1	A_2	A_1	A_2	A_1	A_2
	α^3	α	α^3	β	$\alpha^2\beta$	β	$\alpha\beta\gamma$	α	$\alpha\beta\gamma$	δ
	α	α^3	β	α^3	β	$\alpha^2\beta$	α	$\alpha\beta\gamma$	δ	$\alpha\beta\gamma$
			$\alpha^2\beta$	α	$\alpha\beta^2$	α	$\alpha^2\beta$	γ	$\alpha\beta\delta$	γ
			α	$\alpha^2\beta$	α	$\alpha\beta^2$	γ	$\alpha^2\beta$	γ	$\alpha\beta\delta$
							$\alpha^2\gamma$	β	$\alpha\gamma\delta$	β
							β	$\alpha^2\gamma$	β	$\alpha\gamma\delta$
									$\beta\gamma\delta$	α
									α	$\beta\gamma\delta$
No.	2		4		4		6		8	

in agreement with the theory.

33. Having thus obtained the enumerating function $2h_{w_1}h_{w_2}$ or $h_{w_1}{}^2$ for the special numbers w_1, w_2 we can include all cases by giving w_1, w_2 all possible values and adding the corresponding enumerating functions.

Thus for $w=2$ we have h_1^2,
$\quad$,, $=3$,, $2h_2h_1$,
$\quad$,, $=4$,, $h_2^2+2h_3h_1$,
$\quad$,, $=5$,, $2h_3h_2+2h_4h_1$,

and so on, while in general we seek the coefficient of x^w in the expansion of the function

$$(h_1x+h_2x^2+h_3x^3+\dots)^2.$$

We may state the theorem:—

"The number of distributions of objects of specification $(p_1p_2\dots p_s)$ into boxes of specification (1^2) is equal to the coefficient of the function

$(p_1 p_2 \dots p_s)$ in the development of the coefficient of x^w in the expansion of the function

$$(h_1 x + h_2 x^2 + h_3 x^3 + \dots)^2$$

where $p_1 + p_2 + \dots + p_s = w$."

As an example when $w = 4$,

$$2h_3 h_1 + h_2^2 = 3\,(4) + 6\,(31) + 7\,(2^2) + 10\,(21^2) + 14\,(1^4).$$

The distributions are

Spec.	(4)		(31)		(2^2)		(21^2)		(1^4)	
	A_1	A_2	A_1	A_2	A_1	A_2	A_1	A_2	A_1	A_2
	α^3	α	α^3	β	$\alpha^2\beta$	β	$\alpha^2\beta$	γ	$\alpha\beta\gamma$	δ
	α^2	α^2	β	α^3	β	$\alpha^2\beta$	γ	$\alpha^2\beta$	δ	$\alpha\beta\gamma$
	α	α^3	$\alpha^2\beta$	α	$\alpha\beta^2$	α	$\alpha^2\gamma$	β	$\alpha\beta\delta$	γ
			α	$\alpha^2\beta$	α	$\alpha\beta^2$	β	$\alpha^2\gamma$	γ	$\alpha\beta\delta$
			α^2	$\alpha\beta$	α^2	β^2	$\alpha\beta\gamma$	α	$\alpha\gamma\delta$	β
			$\alpha\beta$	α^2	β^2	α^2	α	$\alpha\beta\gamma$	β	$\alpha\gamma\delta$
					$\alpha\beta$	$\alpha\beta$	α^2	$\beta\gamma$	$\beta\gamma\delta$	α
							$\beta\gamma$	α^2	α	$\beta\gamma\delta$
							$\alpha\beta$	$\alpha\gamma$	$\alpha\beta$	$\gamma\delta$
							$\alpha\gamma$	$\alpha\beta$	$\gamma\delta$	$\alpha\beta$
									$\alpha\gamma$	$\beta\delta$
									$\beta\delta$	$\alpha\gamma$
									$\alpha\delta$	$\beta\gamma$
									$\beta\gamma$	$\alpha\delta$
No.	3		6		7		10		14	

34. Passing to the case of three boxes of specification (1^3) we consider a distribution in which the boxes contain w_1, w_2 and w_3 objects in any order respectively. The possible specifications of these lots of objects are shewn by the partitions of the functions which are terms in $h_{w_1}, h_{w_2}, h_{w_3}$ and when these specifications are assigned the corresponding symmetric function products will be terms of the developed products

$$h_{w_1}{}^3,\ 3h_{w_1}{}^2 h_{w_2},\ 6h_{w_1} h_{w_2} h_{w_3},$$

according to the equalities that present themselves in the numbers

$$w_1,\ w_2,\ w_3.$$

To see how this is we observe that from Art. 27 the distributions associated with specifications

$$(m_1 m_2 \dots m_t),\ (n_1 n_2 \dots n_u),\ (o_1 o_2 \dots o_v)$$

are enumerated by the functions

$$(m_1m_2 \dots m_t)^3, \qquad 3\,(m_1m_2 \dots m_t)^2\,(n_1n_2 \dots n_u),$$
$$6\,(m_1m_2 \dots m_t)\,(n_1n_2 \dots n_u)\,(o_1o_2 \dots o_v),$$

according to the identities that subsist between the three partitions. It is obvious that if $w_1 = w_2 = w_3$ the three partitions are all of the same weight and $h_{w_1}{}^3$ will give the three functions which have coefficients 1, 3, 6 respectively. If w_1, w_1, w_2 be the three weights $3h_{w_1}{}^2h_{w_2}$ involves on development the functions with coefficients 3, 6. Finally if w_1, w_2, w_3 are three different numbers, $6h_{w_1}h_{w_2}h_{w_3}$ produces all the functions which have the coefficient 6.

We can now include all cases by giving w_1, w_2, w_3 all values and adding the corresponding enumerating functions.

Thus for $w = 3$ we have $h_1{}^3$,
4 " $3h_2h_1{}^2$,
5 " $3h_3h_1{}^2 + 3h_1h_2{}^2$,
6 " $h_2{}^3 + 3h_4h_1{}^2 + 6h_3h_2h_1$,
and so on.

In general the enumerating h function is the coefficient of x^w in the expansion of

$$(h_1x + h_2x^2 + h_3x^3 + \dots)^3$$

and if we develop this h function the coefficient of the symmetric function $(p_1p_2 \dots p_s)$ is equal to the number of ways of distributing objects of specification $(p_1p_2 \dots p_s)$ into boxes of specification (1^3).

35. We can now enunciate the *general* theorem :—

"The number of ways of distributing objects of specification $(p_1p_2 \dots p_s)$ into boxes of specification (1^m), no box being empty, is equal to the coefficient of

$$x^{p_1+p_2+\dots+p_s}\,(p_1p_2 \dots p_s)$$

in the development of the function

$$(h_1x + h_2x^2 + h_3x^3 + \dots)^m."$$

In order to be able to use this theorem in practice it is necessary to expand products of the functions h_1, h_2, h_3, ... in terms of monomial functions. This may be readily accomplished by use of the operative symbols D_0, D_1, D_2, ... because observing in the first place that

$$D_0h_3 = D_0\{(3) + (21) + (1^3)\} = (3) + (21) + (1^3) = h_3,$$
$$D_1h_3 = (2) + (1^2) = h_2,$$
$$D_2h_3 = (1) = h_1,$$
$$D_3h_3 = 1,$$

it is easy to see that

$$D_m h_w = h_{w-m}$$

is universally true if we agree that $h_0 = 1$.

If h_w' be the homogeneous product-sum, of weight w, of the quantities $\beta, \gamma, \delta, \dots$ we may write

$$h_w = D_0 h_w' + \alpha D_1 h_w' + \alpha^2 D_2 h_w' + \alpha^3 D_3 h_w' + \dots,$$

so that

$$h_{w_1} h_{w_2} = (D_0 h_{w_1}' + \alpha D_1 h_{w_1}' + \alpha^2 D_2 h_{w_1}' + \dots)(D_0 h_{w_2}' + \alpha D_1 h_{w_2}' + \alpha^2 D_2 h_{w_2}' + \dots).$$

But

$$h_{w_1} h_{w_2} = D_0 (h_{w_1}' h_{w_2}') + \alpha D_1 (h_{w_1}' h_{w_2}') + \alpha^2 D_2 (h_{w_1}' h_{w_2}') + \dots.$$

Hence equating coefficients of like powers of α and suppressing the dashes by writing $\alpha, \beta, \gamma, \dots$ for $\beta, \gamma, \delta \dots$

$$\begin{aligned} D_m (h_{w_1} h_{w_2}) &= D_0 h_{w_1} . D_m h_{w_2} + D_1 h_{w_1} . D_{m-1} h_{w_2} + \dots + D_m h_{w_1} . D_0 h_{w_2} \\ &= h_{w_1} h_{w_2 - m} + h_{w_1 - 1} h_{w_2 - m + 1} + \dots + h_{w_1 - m} h_{w_2}, \end{aligned}$$

shewing the way in which the symbol D_m operates upon any product $h_{w_1} h_{w_2}$. Compare Art. 30.

Similarly D_m operates upon a product of s functions $h_{w_1} h_{w_2} \dots h_{w_s}$ through the medium of the various compositions of m into s parts, zero counting as a part.

36. Thus if we desire to develop the function

$$h_2^3 + 3h_4 h_1^2 + 6h_3 h_2 h_1$$

and require the coefficient of the function (51) the process may be as follows:

$$\begin{aligned} &D_5 (h_2^3 + 3h_4 h_1^2 + 6h_3 h_2 h_1) \\ &= D_2 h_2 . D_2 h_2 . D_1 h_2 + D_2 h_2 . D_1 h_2 . D_2 h_2 + D_1 h_2 . D_2 h_2 . D_2 h_2 \\ &\quad + 3 (D_4 h_4 . D_1 h_1 . D_0 h_1 + D_4 h_4 . D_0 h_1 . D_1 h_1 + D_3 h_4 . D_1 h_1 . D_1 h_1) \\ &\quad + 6 (D_3 h_3 . D_2 h_2 . D_0 h_1 + D_3 h_3 . D_1 h_2 . D_1 h_1 + D_2 h_3 . D_2 h_2 . D_1 h_1) \\ &= 3h_1 + 9h_1 + 18h_1 = 30h_1, \end{aligned}$$

$$D_5 D_1 (h_2^3 + 3h_4 h_1^2 + 6h_3 h_2 h_1) = 30,$$

establishing that objects of specification (51) can be placed in boxes of specification (1^3), no box being empty, in 30 ways.

37. There is an alternative process which is of much interest.

Write $$h_1 x + h_2 x^2 + h_3 x^3 + \dots = H,$$

and note that the coefficient of $x^{p_1 + p_2 + \dots + p_s} (p_1 p_2 \dots p_s)$ in H^m is

$$D_{p_1} D_{p_2} \dots D_{p_s} (\text{coefficient of } x^{p_1 + p_2 + \dots + p_s} \text{ in } H^m).$$

Now $\quad H^m = (1 + H - 1)^m$

$$= (1+H)^m - \binom{m}{1}(1+H)^{m-1} + \binom{m}{2}(1+H)^{m-2} - \dots,$$

and $D_p(1+H) = x^p(1+H)$ by the law of operation.

Also $\quad D_p(1+H)^2 = D_p(1+H) \,.\, D_0(1+H)$
$$+ D_{p-1}(1+H) \,.\, D_1(1+H) + \dots,$$

there being one term on the right-hand side corresponding to every composition of p into two parts.

By Art. 30 the number of these compositions is

$$\binom{p+1}{p} \equiv \binom{p+1}{1}.$$

Hence $\quad x^{-p}D_p(1+H)^2 = \binom{p+1}{1}(1+H)^2;$

also $\quad D_p(1+H)^3 = D_p(1+H) \,.\, D_0(1+H) \,.\, D_0(1+H) + \dots$
$$+ D_a(1+H) \,.\, D_b(1+H) \,.\, D_c(1+H)$$
$$+ \dots,$$

there being one term on the right-hand side corresponding to each composition of p into three parts. The number of these compositions is, by Art. 30, $\binom{p+2}{2}$.

We have $\quad x^{-p}D_p(1+H)^3 = \binom{p+2}{2}(1+H)^3,$

and generally $\quad x^{-p}D_p(1+H)^m = \binom{p+m-1}{m-1}(1+H)^m.$

Making use of these results

$$x^{-p_1}D_{p_1}H^m = \binom{p_1+m-1}{m-1}(1+H)^m$$
$$- \binom{m}{1}\binom{p_1+m-2}{m-2}(1+H)^{m-1}$$
$$+ \binom{m}{2}\binom{p_1+m-3}{m-3}(1+H)^{m-2} + \dots,$$

$$x^{-p_1-p_2}D_{p_1}D_{p_2}H^m = \binom{p_1+m-1}{m-1}\binom{p_2+m-1}{m-1}(1+H)^m$$
$$- \binom{m}{1}\binom{p_1+m-2}{m-2}\binom{p_2+m-2}{m-2}(1+H)^{m-1}$$
$$+ \dots,$$

and ultimately

$$\begin{aligned} x^{-p_1-p_2-\dots-p_s} D_{p_1} D_{p_2} \dots D_{p_s} H^m \\ = \binom{p_1+m-1}{m-1}\binom{p_2+m-1}{m-1}\dots\binom{p_s+m-1}{m-1} \\ - \binom{m}{1}\binom{p_1+m-2}{m-2}\binom{p_2+m-2}{m-2}\dots\binom{p_s+m-2}{m-2} \\ + \binom{m}{2}\binom{p_1+m-3}{m-3}\binom{p_2+m-3}{m-3}\dots\binom{p_s+m-3}{m-3} \\ - \dots, \end{aligned}$$

because we know that the right-hand side cannot involve x. We may therefore finally put x and therefore H equal to zero.

To verify the result of the preceding Article put

$$m=3,\ \ p_1=5,\ \ p_2=1.$$

The formula gives

$$\binom{7}{2}\binom{3}{2}-3\binom{6}{1}\binom{2}{1}+3\binom{5}{0}\binom{1}{0}$$
$$=63-36+3=30.$$

The series written down is thus established as enumerating the number of distributions of objects of specification $(p_1p_2\dots p_s)$ into boxes of specification (1^m), no box being empty.

38. In the above investigation there is no restriction upon the number of times that any one of the quantities $\alpha, \beta, \gamma, \dots$ may appear in the same box.

If no object is to appear more than once in the same box, a box which contains w_1 objects must contain objects denoted by the letters of one of the terms of $a_{w_1} \equiv (1^{w_1})$. Hence instead of the functions $h_1, h_2, h_3, \dots$ we have presented to us the functions $a_1, a_2, a_3, \dots$ and writing

$$a_1x + a_2x^2 + a_3x^3 + \dots = A$$

the enumerating function is the coefficient of $x^{p_1+p_2+\dots+p_s}$ in

$$A^m.$$

If $m=3$, the function which now enumerates the distributions into boxes of specification (1^3) is

$$a_2^3 + 3a_4a_1^2 + 6a_3a_2a_1$$
$$\equiv (1^2)^3 + 3(1^4)(1)^2 + 6(1^3)(1^2)(1),$$

and if the objects be of specification (321) the number of distributions is

$$D_3D_2D_1\{(1^2)^3 + 3(1^4)(1)^2 + 6(1^3)(1^2)(1)\}.$$

By the rule of operation we find

$$D_3\{(1^2)^3+3(1^4)(1)^2+6(1^3)(1^2)(1)\}$$
$$=(1)^3+3(1^3)+6(1^2)(1),$$

D_3D_2 produces $\qquad 3(1)+6(1),$

and finally $\quad D_3D_2D_1\{(1^2)^3+3(1^4)(1)^2+6(1^3)(1^2)(1)\}=9.$

The actual distributions are

A_1	A_2	A_3
$\alpha\beta\gamma$	$\alpha\beta$	α
$\alpha\beta\gamma$	α	$\alpha\beta$
$\alpha\beta$	$\alpha\beta\gamma$	α
$\alpha\beta$	α	$\alpha\beta\gamma$
α	$\alpha\beta\gamma$	$\alpha\beta$
α	$\alpha\beta$	$\alpha\beta\gamma$

A_1	A_2	A_3
$\alpha\beta$	$\alpha\beta$	$\alpha\gamma$
$\alpha\beta$	$\alpha\gamma$	$\alpha\beta$
$\alpha\gamma$	$\alpha\beta$	$\alpha\beta$

39. In the alternative method we write

$$1+A=1+a_1x+a_2x^2+a_3x^3+\dots.$$

The reader will have no difficulty in establishing the formula

$$D_p(1+A)^m=\binom{m}{p}x^p(1+A)^m,$$

so that operating upon A^m in the form

$$(1+A)^m-\binom{m}{1}(1+A)^{m-1}+\binom{m}{2}(1+A)^{m-2}-\dots$$

we readily reach the number which enumerates the distributions of objects of specification $(p_1p_2\dots p_s)$ into boxes of specification (1^m), no box being empty, subject to the condition that no particular object is to appear twice in the same box. The number is

$$\binom{m}{p_1}\binom{m}{p_2}\dots\binom{m}{p_s}-\binom{m}{1}\binom{m-1}{p_1}\binom{m-1}{p_2}\dots\binom{m-1}{p_s}$$
$$+\binom{m}{2}\binom{m-2}{p_1}\binom{m-2}{p_2}\dots\binom{m-2}{p_s}-\dots.$$

To verify the special case $m=3$, $p_1=3$, $p_2=1$, $p_3=1$, we find

$$\binom{3}{3}\binom{3}{2}\binom{3}{1}=9.$$

The more general condition that no object is to appear more than k times in the same box is treated by means of new functions

$$k_1, k_2, k_3, \ldots,$$

such that k_s is derived from h_s by striking out from the latter all partitions which contain parts greater than k. We then operate through the medium of compositions which contain no part greater than k and we reach a general solution analogous to those which employed the h and a functions.

CHAPTER IV

DISTRIBUTION WHEN OBJECTS AND BOXES ARE EQUAL IN NUMBER

40. We now come to an important case of distribution which is of particular interest in view of the light that it throws upon the algebra of symmetric functions. We consider a number of objects and an equal number of boxes. We are given the specifications both of the objects and of the boxes and place one object in each box. How many distributions are there?

Suppose that q_1 of the boxes are precisely similar, so that they have the specification (q_1). Whatever may be the specification of the q_1 objects that are placed in them it is certain that they have only one distribution, because the boxes being identical no permutation of the objects alters the distribution. Denote these boxes each by A_1. The specification of the q_1 objects must be one of the partitions which occur in h_{q_1} when expressed in terms of monomial functions. As one distribution we may take any product of $\alpha, \beta, \gamma, \dots$ that occurs in h_{q_1}. Also if there be q_2 boxes, each denoted by A_2, one distribution into the q_2 boxes will be any product of $\alpha, \beta, \gamma, \dots$ that occurs in h_{q_2}. And similarly for the boxes $q_3, q_4, \dots q_t$. Hence we write down the factors of the product

$$h_{q_1} h_{q_2} \dots h_{q_t},$$

each factor being written out in full, and obtain a distribution by taking any term of h_{q_1} for the q_1 boxes A_1, any term of h_{q_2} for the q_2 boxes A_2, etc. ... any term of h_{q_t} for the q_t boxes A_t. If these terms when assembled constitute a combination which has the specification

$$(p_1 p_2 \dots p_s)$$

we will have one instance of a distribution of objects of specification $(p_1 p_2 \dots p_s)$ into boxes of specification $(q_1 q_2 \dots q_t)$, one object being in each box. It follows that the objects denoted by

$$\alpha_1^{p_1} \alpha_2^{p_2} \dots \alpha_s^{p_s}$$

can be distributed into the boxes just as often as the term $\alpha_1^{p_1} \alpha_2^{p_2} \dots \alpha_s^{p_s}$ arises in the product $h_{q_1} h_{q_2} \dots h_{q_t}$. The enumeration of the distributions is therefore given by the coefficient of the function $(p_1 p_2 \dots p_s)$ in the development of $h_{q_1} h_{q_2} \dots h_{q_t}$ in a series of monomial symmetric functions. We have the theorem:—

"The number of ways of distributing n objects of specification $(p_1 p_2 \ldots p_s)$ into n boxes of specification $(q_1 q_2 \ldots q_t)$, one object into each box, is equal to the coefficient of symmetric function $(p_1 p_2 \ldots p_s)$ in the development of the product $h_{q_1} h_{q_2} \ldots h_{q_t}$."

As an example, if $(p_1 p_2 \ldots p_s) = (411)$, $(q_1 q_2 \ldots q_t) = (321)$, one distribution is

$$\frac{A_1\ A_1\ A_1}{\alpha\ \ \alpha\ \ \alpha} \qquad \frac{A_2\ A_2}{\alpha\ \ \beta} \qquad \frac{A_3}{\gamma}$$

corresponding to the terms α^3, $\alpha\beta$, γ in h_3, h_2, h_1 respectively. The number of distributions is from previous work

$$D_4 D_1^2 h_3 h_2 h_1 = 8,$$

and the complete table of distributions is

A_1	A_1	A_1	A_2	A_2	A_3
α	α	α	α	β	γ
α	α	α	α	γ	β
α	α	α	β	γ	α
α	α	β	α	α	γ
α	α	β	α	γ	α
α	α	γ	α	α	β
α	α	γ	α	β	α
α	β	γ	α	α	α

A table giving the development of products of the functions

$$h_1,\ h_2,\ h_3,\ \ldots$$

will give the complete numerical solution.

41. We now write the particular distribution we presented above in the form, writing A, B, C, ... for A_1, A_2, A_3, ...,

$$\frac{A\ A\ A}{\alpha\ \ \alpha\ \ \alpha} \qquad \frac{B\ B}{\alpha\ \ \beta} \qquad \frac{C}{\gamma}$$

and observe that if we interchange the letters by writing A for α and α for A, B for β and β for B, C for γ and γ for C, we reach a distribution

$$\frac{A\ A\ A\ A}{\alpha\ \ \alpha\ \ \alpha\ \ \beta} \qquad \frac{B}{\beta} \qquad \frac{C}{\gamma}$$

of objects of specification (321) into boxes of specification (411), and

since we may transform every distribution in this way we obtain the theorem :—

" n objects of specification $(p_1 p_2 \dots p_s)$ can be distributed into n boxes of specification $(q_1 q_2 \dots q_t)$, one object in each box, in just as many ways as n objects of specification $(q_1 q_2 \dots q_t)$ can be distributed into n boxes of specification $(p_1 p_2 \dots p_s)$, one object in each box."

42. This quite obvious fact in the Theory of Distributions is next seen to lead to a Theorem of Symmetry in Algebra which is not only not obvious but was for a long time unsuspected.

If we denote by

$$C\begin{pmatrix} p_1 p_2 \dots p_s \\ q_1 q_2 \dots q_t \end{pmatrix}$$

the number of the distributions under examination we have shewn that

$$C\begin{pmatrix} p_1 p_2 \dots p_s \\ q_1 q_2 \dots q_t \end{pmatrix} = C\begin{pmatrix} q_1 q_2 \dots q_t \\ p_1 p_2 \dots p_s \end{pmatrix},$$

and this leads to the relation

$$D_{p_1} D_{p_2} \dots D_{p_s} h_{q_1} h_{q_2} \dots h_{q_t} = D_{q_1} D_{q_2} \dots D_{q_t} h_{p_1} h_{p_2} \dots h_{p_s},$$

or, in other words, the coefficient of symmetric function $(p_1 p_2 \dots p_s)$ in the development of $h_{q_1} h_{q_2} \dots h_{q_t}$ is equal to the coefficient of symmetric function $(q_1 q_2 \dots q_t)$ in the development of $h_{p_1} h_{p_2} \dots h_{p_s}$. This is called a 'Law of Symmetry,' because in a table which expresses the h products in terms of monomials for a given weight the rows will read the same as the columns. Thus such a table for the weight four is

	(4)	(31)	(2^2)	(21^2)	(1^4)
h_4	1	1	1	1	1
$h_3 h_1$	1	2	2	3	4
h_2^2	1	2	3	4	6
$h_2 h_1^2$	1	3	4	7	12
h_1^4	1	4	6	12	24

43. If we look again at the distribution

$$\frac{A\ A\ A}{\alpha\ \ \alpha\ \ \alpha} \qquad \frac{B\ B}{\alpha\ \ \beta} \qquad \frac{C}{\gamma}$$

the symmetry that arises from the interchange of letters leads to the idea that instead of regarding the letters A, B, C as denoting boxes we may regard them as also denoting objects, but of a different kind from

the objects denoted by α, β, γ; so that we may regard the distribution as being in fact a pairing of objects of two different sets of objects, one object being taken from each set to form a pair.

Observe that one set of objects involves no objects which appear in the other set. If the objects of both sets had been drawn from one set of objects, so that the objects in one set were not distinct from the objects in the other set, the distribution theory considered here would not be valid. For example, if we distribute the objects α, β, γ into the boxes A, B, C we obtain the six pairings of the objects α, β, γ with the objects A, B, C,

$$\begin{array}{cccccc} ABC & ABC & ABC & ABC & ABC & ABC \\ \alpha\beta\gamma & \alpha\gamma\beta & \beta\alpha\gamma & \beta\gamma\alpha & \gamma\alpha\beta & \gamma\beta\alpha \end{array}$$

but if we pair off the identical sets α, β, γ; α, β, γ, we obtain only the five pairings

$$\begin{array}{ccccc} \alpha\beta\gamma & \alpha\beta\gamma & \alpha\beta\gamma & \alpha\beta\gamma & \alpha\beta\gamma \\ \alpha\beta\gamma & \alpha\gamma\beta & \beta\alpha\gamma & \beta\gamma\alpha & \gamma\beta\alpha \end{array}$$

because the omitted pairing

$$\begin{array}{c} \alpha\beta\gamma \\ \gamma\alpha\beta \end{array}$$

is the same as

$$\begin{array}{c} \alpha\beta\gamma \\ \beta\gamma\alpha \end{array}$$

When any object in the one set also appears in the other we have a distribution, or pairing, which requires separate consideration, and indeed has been investigated up to a certain point*.

44. The distribution, regarded as a pairing off of sets of objects, which are distinct, is to be regarded as having a specification depending upon similarities of object-pairs. Thus the above pairing may be written

$$(A\alpha)^3 (B\alpha)(B\beta)(C\gamma),$$

which is said to have the specification (3111), which is also a partition of 6, the number of the objects distributed.

We may say that objects of specification (411) have been distributed into boxes of specification (321), one object in each box, in such wise that the specification of the distribution has the specification (3111).

* "Combinations derived from m identical sets of n different letters and their connexion with general magic squares," by Major P. A. MacMahon, *Proc. L. M. S.* Ser. 2, Vol. 17, Part I.

Or, we may say that objects of specification (411) have been paired off with other objects of specification (321) in such wise that the specification of the object-pairs is (3111).

It is next to be noticed that the interchange of Capital and Greek letters does not alter the specification of the distribution. For looking at the object-pairing

$$(A\alpha)^3 (B\alpha)^1 (B\beta)^1 (C\gamma)^1,$$

it is clear that the interchange of letters cannot affect the repetitional numbers 3, 1, 1, 1, which are the parts of the partition which denote the specification.

45. We have before us clearly quite a new question, viz. the enumeration of distributions, of given specification, of objects of specification $(p_1 p_2 \dots p_s)$ into boxes of specification $(q_1 q_2 \dots q_t)$, one object into each box.

In Chapter v this question is considered up to a point. It has been solved completely in Combinatory Analysis. Suffice it to say that the theory has an important bearing upon the Algebra of Symmetric Functions. It establishes a refined law of symmetry connected with the partitions $(p_1 p_2 \dots p_s)$, $(q_1 q_2 \dots q_t)$, and the partition which denotes the distribution due to the circumstance that the first two of these partitions may be interchanged without altering the enumeration.

46. In the present theory the homogeneous product-sums $h_1, h_2, h_3, \dots$ have appeared because no limit was imposed upon the number of times that similar objects may appear in similar boxes. Thus in boxes A, A, A, we have supposed it permissible to place objects represented by any of the terms $\alpha\alpha\alpha$, $\alpha\alpha\beta$, $\alpha\beta\gamma$, ... that compose h_3. The specifications of this portion of the distribution might be (3), (21) or (1^3). If we had resolved that not more than two similar objects were to be placed in similar boxes we could not have placed the objects α, α, α into the boxes A, A, A, and instead of the function h_3 we would have taken the function

$$(21) + (1^3),$$

and generally, in each of the functions h_1, h_2, h_3, ..., we would have deleted all functions which in the partition notation are denoted by partitions which involve parts greater than 2. If the conditions be that not more than k similar objects are to be placed in similar boxes we substitute for $h_1, h_2, h_3, \dots$ the corresponding set of functions $k_1, k_2, k_3, \dots$ in which the deletion of partitions involving parts greater than k has

been carried out. We then find that the number of distributions is equal to the coefficient of the function $(p_1p_2 \dots p_s)$ in the development of

$$k_{q_1}k_{q_2} \dots k_{q_t},$$

and establish by interchange of Capital and Greek letters that the distributions, subject to the same condition, of objects of specification $(q_1q_2 \dots q_t)$ into boxes of specification $(p_1p_2 \dots p_s)$ are enumerated by the same number.

We thus see that the coefficient of $(p_1p_2 \dots p_s)$ in $k_{q_1}k_{q_2} \dots k_{q_t}$ is equal to the coefficient of $(q_1q_2 \dots q_t)$ in $k_{p_1}k_{p_2} \dots k_{p_s}$.

In other words we prove that

$$D_{p_1}D_{p_2} \dots D_{p_s} k_{q_1}k_{q_2} \dots k_{q_t} = D_{q_1}D_{q_2} \dots D_{q_t} k_{p_1}k_{p_2} \dots k_{p_s}.$$

Moreover, since

$$D_p h_m = h_{m-p},$$

also

$$D_p k_m = k_{m-p},$$

the evaluation of the coefficients can be carried out.

The specification of the distribution is clearly not altered by the interchange of Capital and Greek letters and we are led to an extended theory of symmetry in the Algebra of Symmetric Functions.

47. The case $k=1$ is interesting because the homogeneous product-sums become the elementary functions

$$(1), (1^2), (1^3), \dots \equiv a_1, a_2, a_3, \dots,$$

and we establish that the coefficient of the function $(p_1p_2 \dots p_s)$ in the development of

$$a_{q_1}a_{q_2} \dots a_{q_t}$$

is the same as that of $(q_1q_2 \dots q_t)$ in the development of

$$a_{p_1}a_{p_2} \dots a_{p_s}.$$

This particular case of symmetric function symmetry has been known since the time of Meyer Hirsch early in the nineteenth century and several proofs have been given of it. That here given, based upon the theory of distribution, is the simplest and most suggestive. Since the specification of one of these distributions cannot involve any number greater than unity, we see that every distribution must have the same specification, viz. (1^n), where n is the number of objects. In the calculation the symbol D operates entirely through the medium of compositions of numbers which are composed entirely of units and zeros. This is so because

$$D_m(1^p) \equiv D_m a_p = \text{zero if } m \text{ be greater than unity.}$$

Thus

$$D_3 a_2 a_1^2 = D_3(1^2)(1)(1) = D_1(1^2) \,.\, D_1(1) \,.\, D_1(1) = (1),$$

$$D_2(1^2)(1)(1) = D_1(1^2) \,.\, D_1(1) \,.\, D_0(1) + D_1(1^2) \,.\, D_0(1) \,.\, D_1(1) + D_0(1^2) \,.\, D_1(1) \,.\, D_1(1) = 2(1)^2 + (1^2),$$

$$D_1(1^2)(1)(1) = D_1(1^2) \,.\, D_0(1) \,.\, D_0(1) + D_0(1^2) \,.\, D_1(1) \,.\, D_0(1) + D_0(1^2) \,.\, D_0(1) \,.\, D_1(1) = (1)^3 + 2(1^2)(1).$$

48. It has been established that the number

$$D_{p_1} D_{p_2} \ldots D_{p_s} h_{q_1} h_{q_2} \ldots h_{q_t}$$

enumerates distributions of objects into boxes when the distributions are subject to certain conditions.

We can now shew, by reasoning upon the method of obtaining this result, that the same number enumerates certain arithmetical constructions of quite a different nature. When D_{p_1} operates upon $h_{q_1} h_{q_2} \ldots h_{q_t}$ it acts through a number of compositions of p_1 into t parts, zero counting as a part. In this way we obtain the sum of a number of products of which the type is

$$h_{q_1 - c_1} h_{q_2 - c_2} \ldots h_{q_t - c_t},$$

where $c_1 c_2 \ldots c_t$ is a composition of p_1.

Each of these products has unity for coefficient.

Restricting attention to the product above written the operation of D_{p_2} is performed through compositions of p_2, and we obtain from the one product we are attending to a number of products of which the type is

$$h_{q_1 - c_1 - d_1} h_{q_2 - c_2 - d_2} \ldots h_{q_t - c_t - d_t},$$

where $d_1 d_2 \ldots d_t$ is a composition of p_2.

Each of these products has unity for coefficient.

Restricting the attention to this last written product the operation of D_{p_3} yields a number of products of which

$$h_{q_1 - c_1 - d_1 - e_1} h_{q_2 - c_2 - d_2 - e_2} \ldots h_{q_t - c_t - d_t - e_t}$$

is the type, where $e_1 e_2 \ldots e_t$ is a composition of p_3.

Each of these products has unity for coefficient.

Finally, by this process, when we operate with D_{p_s} through one of the compositions of p_s, viz.

$$\sigma_1 \sigma_2 \ldots \sigma_t,$$

we reach the product

$$1 \times h_0 h_0 \ldots h_0 \equiv 1.$$

We will then have arrived at the enumeration of one of our distributions through the medium of the succession of compositions

$$c_1c_2 \dots c_t,\ d_1d_2 \dots d_t,\ e_1e_2 \dots e_t,\ \dots\dots\ \sigma_1\sigma_2 \dots \sigma_t,$$

of the numbers $p_1, p_2, p_3, \dots p_s$ respectively.

We may say that the particular distribution thus enumerated is in correspondence with the numbered diagram

$$\begin{array}{|c|c|c|c|c|}
\hline
c_1 & c_2 & c_3 & \cdots & c_t \\
\hline
d_1 & d_2 & d_3 & \cdots & d_t \\
\hline
e_1 & e_2 & e_3 & \cdots & e_t \\
\hline
\vdots & \vdots & \vdots & & \vdots \\
\hline
\sigma_1 & \sigma_2 & \sigma_3 & \cdots & \sigma_t \\
\hline
\end{array}$$

which involves a rectangle of s rows and t columns.

What is the definition of this diagram? Clearly the sums of the numbers in the successive rows must be $p_1, p_2, p_3, \dots p_s$ respectively, and the sums of the numbers in the successive columns must be $q_1, q_2, q_3, \dots q_t$ respectively. The numbers must be positive integers (zero included) and there is no restriction upon the magnitude.

To every such diagram also there corresponds one distribution. Hence the number

$$D_{p_1}D_{p_2} \dots D_{p_s} h_{q_1}h_{q_2} \dots h_{q_t}$$

enumerates the diagrams so defined.

To take a very simple example, the number

$$D_2^2 h_2 h_1^2 = 4$$

enumerates the diagrams

$$\begin{array}{|c|c|c|}\hline 2 & 0 & 0 \\ \hline 0 & 1 & 1 \\ \hline\end{array} \qquad \begin{array}{|c|c|c|}\hline 1 & 1 & 0 \\ \hline 1 & 0 & 1 \\ \hline\end{array} \qquad \begin{array}{|c|c|c|}\hline 1 & 0 & 1 \\ \hline 1 & 1 & 0 \\ \hline\end{array} \qquad \begin{array}{|c|c|c|}\hline 0 & 1 & 1 \\ \hline 2 & 0 & 0 \\ \hline\end{array}$$

where the rows add up to 2, 2 and the columns to 2, 1, 1.

49. We have an analogous enumeration also when the condition is that not more than k similar objects are to be placed in similar boxes.

In every case the reciprocity that exists between the specifications of the objects and of the boxes can be exhibited by rotating the diagrams through a right angle.

These identities of enumeration are simple instances of a very extensive theory in Combinatory Analysis.

50. Before closing this chapter it may be remarked that the placing of objects of any specification in boxes which are identical, one object in each box, is equivalent from a distribution point of view to placing the same objects in a single box. In both cases the objects can be permuted in any manner without changing the enumeration. There is in fact only one distribution. Consider then a distribution such that q_1 objects are placed in q_1 similar boxes A_1, q_2 objects in q_2 similar boxes A_2, ... q_t objects in similar boxes A_t, the sets of objects having any specifications and one object being in each box. In contrast with this consider the $q_1, q_2, \dots q_t$ objects placed in *single* boxes $B_1, B_2, \dots B_t$ respectively. If the numbers $q_1, q_2, \dots q_t$ be all different we cannot in the first distribution interchange any pair of the sets of $q_1, q_2, \dots q_t$ objects because, for example, the q_r objects will only fit into the q_r similar boxes A_r. Also in the second distribution if the boxes $B_1, B_2, \dots B_t$ be identical we cannot alter the distribution by any interchange of a pair of the sets of $q_1, q_2, \dots q_t$ objects. Hence there is a one-to-one correspondeuce and we may state that the number of distributions of objects of specification $(p_1 p_2 \dots p_s)$ into boxes of specification $(q_1 q_2 \dots q_t)$, *the numbers* $q_1, q_2, \dots q_t$ *being all different*, one objec being placed in each box, is equal to the number of distributions of objects of specification $(p_1 p_2 \dots p_s)$ into boxes of specification (t) such that the t boxes contain $q_1, q_2, \dots q_t$ objects respectively. For example, compare these distributions where

$$(p_1 p_2 \dots p_s) = (321), \quad (q_1 q_2 \dots q_t) = (321),$$

$\underline{A_1 A_1 A_1}$	$\underline{A_2 A_2}$	$\underline{A_3}$
	or	
$\underline{A}$	$\underline{A}$	$\underline{A}$
$\alpha\beta\gamma$	$\alpha\beta$	α
$\alpha\beta\beta$	$\alpha\gamma$	α
$\alpha\alpha\gamma$	$\beta\beta$	α
$\beta\beta\gamma$	$\alpha\alpha$	α
$\alpha\alpha\beta$	$\beta\gamma$	α
$\alpha\beta\gamma$	$\alpha\alpha$	β
$\alpha\alpha\gamma$	$\alpha\beta$	β
$\alpha\alpha\beta$	$\alpha\gamma$	β
$\alpha\alpha\alpha$	$\beta\gamma$	β
$\alpha\alpha\alpha$	$\beta\beta$	γ
$\alpha\alpha\beta$	$\alpha\beta$	γ
$\alpha\beta\beta$	$\alpha\alpha$	γ

51. Again, if the numbers $q_1, q_2, \dots q_t$ be identical and the boxes $B_1, B_2, \dots B_t$ have the specification (1^t) we find that in the first distribution the sets of $q_1, q_2, \dots q_t$ objects can be permuted in all possible ways so as to produce new distributions—the number of ways depending upon the similarities that may exist between the t sets of objects. Also in the second distribution, since the boxes are all different, the sets of objects can be permuted exactly as in the first distribution, and we may say that the number of distributions of objects of specification $(p_1 p_2 \dots p_s)$ into boxes of specification $(q_1 q_2 \dots q_t)$, *the numbers* $q_1, q_2, \dots q_t$ *being identical*, one object being placed in each box, is equal to the number of distributions of objects of specification $(p_1 p_2 \dots p_s)$ into boxes of specification (1^t) such that the t boxes contain in some order $q_1, q_2, \dots q_t$ objects respectively. As an example we may compare the distributions of objects of specification (321) into the boxes

$$A_1 A_1 \;\; A_2 A_2 \;\; A_3 A_3 \text{ and } A_1 A_2 A_3,$$

where $(p_1 p_2 \dots p_s) = (321)$, $(q_1 q_2 \dots q_t) = (222)$.

CHAPTER V

DISTRIBUTIONS OF GIVEN SPECIFICATION

52. In this chapter we examine the distribution theory that has just been before us with special reference to the specifications of the distributions. In a product-sum such as h_3, for example

$$(3)+(21)+(1^3),$$

the occurrence of a part 1, 2, or 3 in a partition indicates that 1, 2, or 3 similar parts have been placed in similar boxes and it was by restricting the magnitude of these parts to be not greater than k that we were able to determine the theory of the distribution when the condition was that not more than k similar objects were to be placed in similar boxes. In order to put in evidence the specifications of the distributions we consider in connexion with the product-sums $h_1, h_2, h_3, \ldots$ the new functions

$$\begin{aligned}
X_1 &= x_1(1),\\
X_2 &= x_2(2)+x_1^2(1^2),\\
X_3 &= x_3(3)+x_2x_1(21)+x_1^3(1^3),\\
X_4 &= x_4(4)+x_3x_1(31)+x_2^2(2^2)+x_2x_1^2(21^2)+x_1^4(1^4).
\end{aligned}$$

We may if we choose regard $x_1, x_2, x_3, \ldots$ as being the elementary symmetric functions of a new set of elements

$$\alpha', \beta', \gamma', \ldots.$$

Indicating symmetric functions of this set by dashed brackets, viz. $(\quad)'$, the relations may be written

$$\begin{aligned}
X_1 &= (1)'(1),\\
X_2 &= (1^2)'(2)+(2)'(1^2)+2(1^2)'(1^2),\\
X_3 &= (1^3)'(3)+2(1^3)'(21)+(21)'(21)+6(1^3)'(1^3)\\
&\quad +(3)'(1^3)+2(21)'(1^3),\\
&\quad \text{etc.}
\end{aligned}$$

and we, at once, notice a symmetry in the right-hand sides of these relations. They are unaltered by an interchange of dashed and undashed brackets or in other words, by an interchange of the sets of quantities $\alpha, \beta, \gamma, \ldots \alpha', \beta', \gamma', \ldots$. To prove that this symmetry is universal consider the infinite series

$$1+X_1+X_2+X_3+\ldots,$$

which is expressible as the product

$$\begin{aligned}&(1 + x_1\alpha + x_2\alpha^2 + x_3\alpha^3 + \ldots)\\ \times\,&(1 + x_1\beta + x_2\beta^2 + x_3\beta^3 + \ldots)\\ \times\,&(1 + x_1\gamma + x_2\gamma^2 + x_3\gamma^3 + \ldots)\\ \times\,&\ldots,\end{aligned}$$

because the coefficient of $x_\lambda x_\mu x_\nu \ldots$ therein is

$$\Sigma\, \alpha^\lambda \beta^\mu \gamma^\nu \ldots \equiv (\lambda\mu\nu\ldots).$$

Since $x_1, x_2, x_3, \ldots$ are the elementary functions of $\alpha', \beta', \gamma', \ldots$

$$(1 + x_1\alpha + x_2\alpha^2 + x_3\alpha^3 + \ldots) = (1 + \alpha'\alpha)(1 + \beta'\alpha)(1 + \gamma'\alpha)\ldots,$$

so that also

$$\begin{aligned}&1 + X_1 + X_2 + X_3 + \ldots\\ =\,&(1 + \alpha'\alpha)(1 + \beta'\alpha)(1 + \gamma'\alpha)\ldots\\ \times\,&(1 + \alpha'\beta)(1 + \beta'\beta)(1 + \gamma'\beta)\ldots\\ \times\,&(1 + \alpha'\gamma)(1 + \beta'\gamma)(1 + \gamma'\gamma)\ldots\\ \times\,&\ldots;\end{aligned}$$

a relation which establishes the symmetry for the right-hand side is unaltered by the interchange of dashed and undashed letters.

53. We have to deal at present with the set of relations which commences with $X_1 = x_1(1)$.

Taking any product of the functions X, say for example X_4X_3, we find that we can arrange the right-hand side according to products of quantities $x_1, x_2, x_3, \ldots$. In particular, selecting the term which involves $x_3x_2x_1^2$, we have

$$X_4X_3 = \ldots + \{(21^2)(3) + (31)(21)\}\, x_3x_2x_1^2 + \ldots.$$

The function

$$(21^2)(3) + (31)(21)$$

is associated with two partitions of the number 7; (43) which defines the X product and (321^2) which defines the x product. The numbers which appear in the two functions $(21^2)(3)$, $(31)(21)$ are those which appear in the x product and moreover each function involves partitions of the numbers 4, 3 which appear in the X product.

The symmetric function products $(21^2)(3)$, $(31)(21)$ are derived from the symmetric function (321^2) by a process called 'Separation' and each is said to be a 'Separation' of (321^2). Each factor of such a product is said to be a 'Separate' of the 'Separation.' The like terms are employed when we are thinking only of Partitions. A partition is separated into separates just as a number is partitioned into parts.

Separation consists in separating combinations of parts by distinct brackets. Thus

$$(321^2),\ (321)(1),\ (31^2)(2),\ (32)(1^2),\ (21^2)(3),\ (31)(21),\ (32)(1)^2,$$
$$(31)(2)(1),\ (3)(21)(1),\ (3)(2)(1^2),$$
$$(3)(2)(1)^2,$$

are all separations alike of the function (321^2) and of the partition (321^2).

We may therefore say that the two terms of $(21^2)(3)+(31)(21)$ are, both, separations of the function (321^2).

A separation has a specification which consists of the series of numbers which denote the sums of the numbers in separate brackets or as we may say in the separates. Thus the eleven separations above written have specifications

$$(7),\ (61),\ (52),\ (52),\ (43),\ (43),\ (51^2),$$
$$(421),\ (3^21),\ (32^2),$$
$$(321^2).$$

Hence the terms of $(21^2)(3)+(31)(21)$ may be fully described as being separations, of the partition (321^2) which defines the x product, which have the specification (43) which defines the X product. The terms $(21^2)(3)$, $(31)(21)$ each appear above with the coefficient unity because in the associated X product no exponent exceeds unity. Had we chosen the product $X_3^2X_2^3$ we would have obtained a term

$$2(3)(21)\,.\,3(2)^2(1^2)\ \ x_3x_2^3x_1^3$$

such that $(3)(21)(2)^2(1^2)$ is a separation of (32^31^3) of specification (3^22^3) and the coefficient 3×2 that presents itself denotes that the separation $(2)^2(1^2)$, composed of separates of the same weight, has three permutations; and similarly that the separation $(3)(21)$, also composed of separates of the same weight, has two permutations.

We may say that in the X product the coefficient of a separation is equal to the number of permutations of the separates when only permutations between separates of the same weight are permitted.

54. Take now the general X product

$$X_{q_1}X_{q_2}\dots X_{q_t} = \dots + P\,x_{\sigma_1}x_{\sigma_2}x_{\sigma_3}\dots + \dots.$$

We see that

(i) P is a linear function of separations of $(\sigma_1\sigma_2\sigma_3\dots)$.

(ii) Each separation that appears has the specification $(q_1q_2\dots q_t)$ and every such separation presents itself.

(iii) The numerical coefficient of a separation is equal to the number of permutations of its separates when only permutations between separates of the same weight are permitted.

We now expand P in a series of monomials so that

$$P = \ldots + \theta(p_1 p_2 \ldots p_s) + \ldots$$

and

$$X_{q_1} X_{q_2} \ldots X_{q_t} = \ldots + \theta(p_1 p_2 \ldots p_s)\, x_{\sigma_1} x_{\sigma_2} x_{\sigma_3} \ldots.$$

We gather that objects of specification

$$(p_1 p_2 \ldots p_s)$$

can be distributed into boxes of specification

$$(q_1 q_2 \ldots q_t),$$

one object in each box, so that the distributions have, all, the specification

$$(\sigma_1 \sigma_2 \sigma_3 \ldots)$$

in θ ways.

55. It has been seen in the foregoing chapter that we can interchange the specifications of the objects and boxes without altering the specifications of the distributions or the number θ. Hence we have a law of algebraic symmetry indicated by the complementary formula

$$X_{p_1} X_{p_2} \ldots X_{p_s} = \ldots + \theta(q_1 q_2 \ldots q_t)\, x_{\sigma_1} x_{\sigma_2} x_{\sigma_3} \ldots + \ldots.$$

As an example we develop the term

$$\{(21^2)(3) + (31)(21)\}\, x_3 x_2 x_1^2,$$

which appears in the product $X_4 X_3$, and we find

$$X_4 X_3 = \ldots + \{(52) + 3(51^2) + (43) + 2(421) + 2(3^2 1) + 2(32^2) + 3(321^2)\}\, x_3 x_2 x_1^2 + \ldots$$

and we interpret any particular term, say $3(51^2)$, by stating that objects of specification (51^2) can be distributed into boxes of specification (43), one object in each box, in such wise that the distribution has a specification (321^2) in 3 ways. These are in fact

A	A	A	A	B	B	B
α	α	α	β	α	α	γ
α	α	α	γ	α	α	β
α	α	β	γ	α	α	α

the specifications of the distributions being shewn by

$$\begin{array}{ccc|cc|c|c} A & A & A & B & B & A & B \\ \hline \alpha & \alpha & \alpha & \alpha & \alpha & \beta & \gamma \end{array} \qquad \begin{array}{ccc|cc|c|c} A & A & A & B & B & A & B \\ \hline \alpha & \alpha & \alpha & \alpha & \alpha & \gamma & \beta \end{array}$$

$$\begin{array}{ccc|cc|c|c} B & B & B & A & A & A & A \\ \hline \alpha & \alpha & \alpha & \alpha & \alpha & \beta & \gamma \end{array}$$

56. To shew the reciprocity we calculate

$$X_5 X_1^2 = \ldots + 3\,(43)\,x_3 x_2 x_1^2 + \ldots,$$

and the distributions are

$$\begin{array}{ccccccc} A & A & A & A & A & B & C \\ \hline \alpha & \alpha & \alpha & \beta & \beta & \alpha & \beta \\ \alpha & \alpha & \alpha & \beta & \beta & \beta & \alpha \\ \alpha & \alpha & \beta & \beta & \beta & \alpha & \alpha \end{array}$$

3 in number and each of specification (321^2).

57. The symbol D_m can be employed with good effect because if we operate upon

$$X_{q_1} X_{q_2} \ldots X_{q_t}$$

with

$$D_{p_1} D_{p_2} \ldots D_{p_s},$$

we obtain a linear function of x products which gives a complete specification account of the distributions of the objects into the boxes.

We proceed from the relation

$$D_s X_q = x_s X_{q-s},$$

valid for all integer values of s and also when $s = 0$ if we put $x_0 = 1$.

To take an example consider objects and boxes of the specifications $(2^2 1^3)$, (43) respectively and recall the way in which the symbol D_m operates upon a product through the compositions of its suffix. The calculation is

$$\begin{aligned} & D_2^2 D_1^3 X_4 X_3 \\ &= D_2 D_1^3 (x_2 X_3 X_2 + x_2 X_4 X_1 + x_1^2 X_3 X_2) \\ &= D_1^3 \{(x_2 + x_1^2)(x_2 X_3 + x_2 X_2 X_1 + x_1^2 X_2 X_1) \\ &\qquad\qquad + x_2 (x_2 X_2 X_1 + x_1^2 X_3)\} \\ &= D_1^3 \{(x_2^2 + 2x_2 x_1^2) X_3 + (2x_2^2 + 2x_2 x_1^2 + x_1^4) X_2 X_1\} \\ &= D_1^2 \{(x_2^2 x_1 + 2x_2 x_1^3) X_2 + (2x_2^2 x_1 + 2x_2 x_1^3 + x_1^5) X_2 \\ &\qquad\qquad + (2x_2^2 x_1 + 2x_2 x_1^3 + x_1^5) X_1^2\} \\ &= D_1^2 \{(3x_2^2 x_1 + 4x_2 x_1^3 + x_1^5) X_2 + (2x_2^2 x_1 + 2x_2 x_1^3 + x_1^5) X_1^2\} \\ &= D_1 (7x_2^2 x_1^2 + 8x_2 x_1^4 + 3x_1^6) X_1 \\ &= 7x_2^2 x_1^3 + 8x_2 x_1^5 + 3x_1^7, \end{aligned}$$

and the distributions indicated are

Spec.	$(2^2 1^3)$		(21^5)		(1^7)	
	$A\,A\,A\,A$	$B\,B\,B$	$A\,A\,A\,A$	$B\,B\,B$	$A\,A\,A\,A$	$B\,B\,B$
	$\alpha\ \alpha\ \beta\ \beta$	$\gamma\ \delta\ \epsilon$	$\alpha\ \alpha\ \beta\ \gamma$	$\beta\ \delta\ \epsilon$	$\alpha\ \beta\ \gamma\ \delta$	$\alpha\ \beta\ \epsilon$
	$\alpha\ \alpha\ \gamma\ \delta$	$\beta\ \beta\ \epsilon$	$\alpha\ \alpha\ \beta\ \delta$	$\beta\ \gamma\ \epsilon$	$\alpha\ \beta\ \gamma\ \epsilon$	$\alpha\ \beta\ \delta$
	$\alpha\ \alpha\ \gamma\ \epsilon$	$\beta\ \beta\ \delta$	$\alpha\ \alpha\ \beta\ \epsilon$	$\beta\ \gamma\ \delta$	$\alpha\ \beta\ \delta\ \epsilon$	$\alpha\ \beta\ \gamma$
	$\alpha\ \alpha\ \delta\ \epsilon$	$\beta\ \beta\ \gamma$	$\alpha\ \gamma\ \beta\ \beta$	$\alpha\ \delta\ \epsilon$		
	$\gamma\ \delta\ \beta\ \beta$	$\alpha\ \alpha\ \epsilon$	$\alpha\ \delta\ \beta\ \beta$	$\alpha\ \gamma\ \epsilon$		
	$\gamma\ \epsilon\ \beta\ \beta$	$\alpha\ \alpha\ \delta$	$\alpha\ \epsilon\ \beta\ \beta$	$\alpha\ \gamma\ \delta$		
	$\delta\ \epsilon\ \beta\ \beta$	$\alpha\ \alpha\ \gamma$	$\alpha\ \gamma\ \delta\ \epsilon$	$\alpha\ \beta\ \beta$		
			$\beta\ \gamma\ \delta\ \epsilon$	$\alpha\ \alpha\ \beta$		
No.	7		8		3	

58. It will be observed that, since $D_s X_q = x_s X_{q-s}$, the x product of highest degree obtained from

$$D_{p_1} D_{p_2} \dots D_{p_s} X_{q_1} X_{q_2} \dots X_{q_t}$$

must be

$$x_{p_1} x_{p_2} \dots x_{p_s}.$$

Again from the symmetry on the right-hand sides of the relations

$$X_1 = x_1(1),$$
$$X_2 = x_2(2) + x_1^2(1^2),$$
$$X_3 = x_3(3) + x_2 x_1(21) + x_1^3(1^3),$$

which was established in Art. 52, we may derive from the relation

$$D_s X_q = x_s X_{q-s}$$

the relation

$$D_s' X_q = a_s X_{q-s}$$

where the symbol D_s' has reference to the symmetric functions of the quantities $\alpha', \beta', \gamma', \dots$ and as before $a_1, a_2, a_3, \dots$ are the elementary functions of the quantities $\alpha, \beta, \gamma, \dots$.

This is so because an interchange of the sets

$$\alpha, \beta, \gamma, \dots, \qquad \alpha', \beta', \gamma', \dots$$

leaves X_q and X_{q-s} unaltered while changing D_s into D_s' and x_s into a_s.

Similarly from the result

$$D_2^2 D_1^3 X_4 X_3 = 7x_1^3 x_2^2 + 8x_1^5 x_2 + 3x_1^7,$$

we derive

$$D_2'^2 D_1'^3 X_4 X_3 = 7a_1^3 a_2^2 + 8a_1^5 a_2 + 3a_1^7.$$

These transformations are of much service in the development of the algebra.

59. In Art. 55 we have determined the specifications of the distributions when we are given the specifications of the objects and boxes. We can obtain all the distributions which have a given specification, the specifications of the objects and boxes being at disposal by simply expanding an X product as a linear function of x products. Thus since

$$X_2^2 = x_2^2\{(4) + 2(2^2)\} + x_2x_1^2\{2(31) + 2(21^2)\} + x_1^4\{(2^2) + 2(21^2) + 6(1^4)\}$$

we gather that a distribution of specification (2^2) can be obtained, when the box specification is (2^2), by distributing objects of specification (4) in one way, and objects of specification (2^2) in two ways; and similarly the other two terms upon the right-hand side can be interpreted.

The distributions are

$$\begin{array}{c|cc|cc}
\multicolumn{5}{c}{(2^2)} \\
 & A & A & B & B \\
\hline
(4) & \alpha & \alpha & \alpha & \alpha \\
\hline
(2^2) & \alpha & \alpha & \beta & \beta \\
 & \beta & \beta & \alpha & \alpha
\end{array}
\qquad
\begin{array}{c|cc|cc}
\multicolumn{5}{c}{(21^2)} \\
 & A & A & B & B \\
\hline
(31) & \alpha & \alpha & \alpha & \beta \\
 & \alpha & \beta & \alpha & \alpha \\
\hline
(21^2) & \alpha & \alpha & \beta & \gamma \\
 & \beta & \gamma & \alpha & \alpha
\end{array}
\qquad
\begin{array}{c|cc|cc}
\multicolumn{5}{c}{(1^4)} \\
 & A & A & B & B \\
\hline
(2^2) & \alpha & \beta & \alpha & \beta \\
\hline
(21^2) & \alpha & \beta & \alpha & \gamma \\
 & \alpha & \gamma & \alpha & \beta \\
\hline
 & \alpha & \beta & \gamma & \delta \\
 & \alpha & \gamma & \beta & \delta \\
(1^4) & \alpha & \delta & \beta & \gamma \\
 & \beta & \gamma & \alpha & \delta \\
 & \beta & \delta & \alpha & \gamma \\
 & \gamma & \delta & \alpha & \beta
\end{array}$$

60. In Art. 48 we shewed that the theory of a certain distribution led easily to the enumeration of certain numbered diagrams which could be accurately defined. The correspondence was obtained by an examination of the way in which the operation of the symbol D_m is effective in obtaining the enumerating number. Looking back to Art. 57 we can similarly examine the calculation involved in the expression

$$D_2^2 D_1^3 X_4 X_3.$$

The symbol D_m is performed through the medium of the compositions of the number m. If

$$c_1, c_2, \ldots c_t$$

be such a composition we may have to perform the symbols

$$D_{c_1}, D_{c_2}, D_{c_3}, \ldots D_{c_t}$$

upon the several factors of the X product. Now since (Art. 57)

$$D_s X_q = x_s X_{q-s}$$

we see that, associated with the particular portion of the operation, we will have a product

$$x_{c_1}\, x_{c_2} \ldots x_{c_t}$$

with coefficient unity, and not merely unity as is the case when we are dealing with the functions $h_1, h_2, h_3, \ldots$.

Again operating as in Art. 48 through another composition $d_1, d_2, \ldots d_t$ we obtain another x factor

$$x_{d_1}\, x_{d_2} \ldots x_{d_t}$$

the coefficient being again unity.

Finally we arrive at a certain x product with the coefficient unity and we find that corresponding to one of the distributions we have a lettered diagram

$$\begin{array}{|c|c|c|c|c|}
\hline
x_{c_1} & x_{c_2} & x_{c_3} & \cdots & x_{c_t} \\
\hline
x_{d_1} & x_{d_2} & x_{d_3} & \cdots & x_{d_t} \\
\hline
x_{e_1} & x_{e_2} & x_{e_3} & \cdots & x_{e_t} \\
\hline
\vdots & \vdots & \vdots & & \vdots \\
\hline
x_{s_1} & x_{s_2} & x_{s_3} & \cdots & x_{s_t} \\
\hline
\end{array}$$

and the product of these st, x factors defines the specification of the particular distribution which has led us to this diagram. Eliminating the symbol x from the diagram, as it has no numerical significance, we write it

$$\begin{array}{|c|c|c|c|c|}
\hline
c_1 & c_2 & c_3 & \cdots & c_t \\
\hline
d_1 & d_2 & d_3 & \cdots & d_t \\
\hline
e_1 & e_2 & e_3 & \cdots & e_t \\
\hline
\vdots & \vdots & \vdots & & \vdots \\
\hline
s_1 & s_2 & s_3 & \cdots & s_t \\
\hline
\end{array}$$

If the distributions considered have, all of them, the same specification defined by the x product resulting from the diagram, it is clear that the diagrams must all give the product $x_{\sigma_1} x_{\sigma_2} x_{\sigma_3} \ldots$, where the

partition $(\sigma_1\sigma_2\sigma_3 \dots)$ is the specification of the distributions. Hence the numbers $c_1, c_2, \dots d_1, d_2, \dots s_1, s_2, \dots$ in the numbered diagrams must, when assembled, make up the partition $(\sigma_1\sigma_2\sigma_3 \dots)$.

We may therefore state the theorem :—

"Let there be a rectangular chess-board of t columns and s rows and let given numbers $\sigma_1, \sigma_2, \dots \sigma_\tau$, τ in number, be placed in the compartments in such wise that the sums of the numbers in the successive rows are $p_1, p_2, \dots p_s$ and in the successive columns $q_1, q_2, \dots q_t$; then the number of such numbered chess-boards is equal to the number of ways of distributing objects of specification $(p_1p_2 \dots p_s)$ into boxes of specification $(q_1q_2 \dots q_t)$ subject to the condition that the distributions are to have the specification $(\sigma_1\sigma_2 \dots \sigma_t)$."

61. As an example, by means of the result

$$D_2^2 D_1^3 X_4 X_3 = 7x_1^3x_2^2 + 8x_1^5x_2 + 3x_1^7,$$

we enumerate the 7 diagrams

2		2		2		2			2		2		2
2			2		2		2	2		2		2	
	1	1		1			1	1		1			1
	1	1			1	1		1			1	1	
	1		1	1		1			1	1		1	

in which the compartment numbers are 2, 2, 1, 1, 1 derived from the specification (2^21^3) of the corresponding distributions; the row sums are 2, 2, 1, 1, 1 derived from the specification (2^21^3) of the objects distributed, and the column sums are 4, 3 derived from the specification (43) of the boxes.

Also we enumerate the 8 diagrams

2		2		2		1	1	1	1	1	1	1	1		2
1	1	1	1	1	1	2		2		2			2	1	1
1			1		1	1			1		1	1		1	
	1	1			1		1	1			1	1		1	
	1		1	1			1		1	1		1		1	

associated with the compartment numbers 2, 1, 1, 1, 1, 1, and the 3 diagrams

1	1
1	1
1	
1	
	1

1	1
1	1
1	
	1
1	

1	1
1	1
	1
1	
1	

associated with the compartment numbers 1, 1, 1, 1, 1, 1, 1, the row and column sums being, as before, derived from the partitions $(2^2 1^3)$, (43). (Compare Art. 49.)

CHAPTER VI

THE MOST GENERAL CASE OF DISTRIBUTION

62. So far two main divisions of the Theory of Distribution have been under consideration, viz. the case in which there are no similarities between the boxes and the case in which the number of boxes is equal to the number of objects. In the former the objects may be of any specification; in the latter both the objects and the boxes may be of any specification. The next main division that presents itself for examination is concerned with boxes which may be any in number but in every case indistinguishable from one another. They have the specification (m) when they are m in number. The objects may be any in number and of any specification. No box is supposed to be left empty so that the objects are at least as numerous as the boxes.

Objects of the specification $(p_1 p_2 \ldots p_s)$ are in correspondence with the assemblage of letters

$$\alpha_1^{p_1}\alpha_2^{p_2}\ldots\alpha_s^{p_s} \text{ or } \alpha^{p_1}\beta^{p_2}\ldots\sigma^{p_s} \text{ or } \alpha^p\beta^q\gamma^r\ldots.$$

63. The partition $(p_1 p_2 \ldots p_s)$ may, from another standpoint, be regarded as a multipartite number or, in other words, as a succession of numbers which enumerate letters or objects of different kinds.

If we separate any combination of letters from the assemblage

$$\alpha^p\beta^q\gamma^r\ldots,$$

say

$$\alpha^{p_1}\beta^{q_1}\gamma^{r_1}\ldots,$$

the numbers p_1, q_1, r_1, ... are not necessarily or generally in descending order of magnitude and some of them may be zeros. If we break up the assemblage into m portions

$$\alpha^{p_1}\beta^{q_1}\gamma^{r_1}\ldots,\ \alpha^{p_2}\beta^{q_2}\gamma^{r_2}\ldots,\ \ldots\ \alpha^{p_m}\beta^{q_m}\gamma^{r_m}\ldots,$$

without any regard to the order of writing the portions, we may speak of a distribution of objects of specification $(pqr\ldots)$ into boxes of specification (m) because no permutation of the boxes, which are all similar, alters the distribution. In correspondence we speak of partitioning the multipartite number into m multipartite parts and we denote such partition by the notation

$$(p_1q_1r_1\ldots,\ p_2q_2r_2\ldots,\ \ldots p_mq_mr_m\ldots).$$

The parts may be placed in any order without affecting the partition.

Thence it arises that the problem of distribution into similar boxes is identical with that of partitioning a multipartite number.

It will be remarked that a collection of integers in a bracket may denote either a partition of an ordinary or unipartite number or a multipartite number, but that whereas the parts of the partition in the former case may always be written in descending order, such is not the case with the constituents of the multipartite parts of a multipartite number.

As a simple example of the correspondence between distribution and partition, take the assemblage $\alpha^2\beta^2$.

Distribution of $\alpha^2\beta^2$ into two similar boxes		Partitions of (22) into two parts
A	A	
$\alpha^2\beta$	β	(21, 01)
$\alpha\beta^2$	α	(12, 10)
α^2	β^2	(20, 02)
$\alpha\beta$	$\alpha\beta$	(11, 11)

64. In the main divisions previously discussed we have had to deal with the homogeneous product-sums of the elements $\alpha, \beta, \gamma, \ldots$. In the present main division we have also to deal with homogeneous product-sums, not of the simple elements but of certain combinations of them, $u_1, u_2, u_3, \ldots$.

A reference to Art. 8 shews that we can arrive at the product-sums by first obtaining the power-sums.

Thus if
$$u_1^k + u_2^k + u_3^k + \ldots = \sigma_k,$$
and $U_1, U_2, U_3, \ldots$ denote the product-sums,

$$U_1 = \sigma_1,$$
$$2!\,U_2 = \sigma_1^2 + \sigma_2,$$
$$3!\,U_3 = \sigma_1^3 + 3\sigma_1\sigma_2 + 2\sigma_3,$$
$$4!\,U_4 = \sigma_1^4 + 6\sigma_1^2\sigma_2 + 3\sigma_2^2 + 8\sigma_1\sigma_3 + 6\sigma_4,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$m!\,U_m = \Sigma \frac{m!}{m_1!\,m_2!\,m_3!\ldots}\left(\frac{\sigma_1}{1}\right)^{m_1}\left(\frac{\sigma_2}{2}\right)^{m_2}\left(\frac{\sigma_3}{3}\right)^{m_3}\ldots.$$

We have to determine the particular combinations of $\alpha, \beta, \gamma, \ldots$ that we may substitute for $u_1, u_2, u_3, \ldots$, so as to be of service in the problem before us.

If we take $m = 1$, so that there is but a single box, we note that for any assemblage of objects

$$\alpha^p\beta^q\gamma^r\ldots$$

there is only one distribution; the whole of the objects must be placed in the only box. Hence the symmetric-function enumerating function must be the sum of all the monomial functions of all weights. We may take it to be

$$h_1 + h_2 + h_3 + \ldots \text{ ad inf.}$$
$$\equiv (1) + (2) + (1^2) + (3) + (21) + (1^3) + \ldots,$$

because the coefficient of the function $(pqr\ldots)$ in the series of functions is unity.

When $m = 2$, we may place in the two boxes any two assemblages which, added together, make the assemblage to be distributed. Regarding

$$\alpha^p\beta^q\gamma^r\ldots$$

as a literal product, we have two products whose product is equal to the given product. Now it is evident that if P_1, P_2 be two such products, the distribution must be of one of the *types*

$$\frac{A\ A}{P_1P_1} \quad \frac{A\ A}{P_1P_2}$$

so that the product distributed must be either P_1^2 or P_1P_2, where P_1, P_2 separately may be any combination of letters. Hence every possible distribution will be realised for all specifications of the objects to be distributed by taking the product-sums of *order two* of all combinations of letters. The enumerating function must therefore be the sum of such product-sums of all weights.

Similarly when $m = 3$, the distribution must be of one of the types

$$\frac{A\ A\ A}{P_1P_1P_1} \quad \frac{A\ A\ A}{P_1P_1P_2} \quad \frac{A\ A\ A}{P_1P_2P_3}$$

so that the product to be distributed must be either P_1^3 or $P_1^2P_2$ or $P_1P_2P_3$. Hence the enumerating function must be the sum of product-sums of *order three* of all combinations of letters.

By similar reasoning for m boxes the enumerating function must be the sum of product-sums of *order* m of all combinations of letters. These combinations are the terms of the infinite series

$$h_1 + h_2 + h_3 + \ldots \text{ ad. inf.}$$

65. If we proceed now from these combinations we will obtain a solution of the problem, but it is much better to include unity in the series of terms. If unity may be placed in any box instead of one of the above combinations it is clear that we will enumerate the distributions into m *or fewer* boxes, and this will be quite satisfactory because we have only to subtract the function which enumerates the distributions into $m-1$ or fewer boxes in order to obtain the function which enumerates the distributions into m boxes, no box being empty. As the algebra is easier we adopt this course and put

$$S_1 = 1 + h_1 + h_2 + h_3 + \dots$$
$$\equiv 1 + (1) + (2) + (1^2) + (3) + (21) + (1^3) + \dots.$$

The sum of the kth powers S_k, of all the terms appearing herein, is obtained, as is readily realised, by multiplying every part which appears in the partitions by k.

Hence

$$S_2 = 1 + (2) + (4) + (2^2) + (6) + (42) + (2^3) + \dots,$$
$$S_3 = 1 + (3) + (6) + (3^2) + (9) + (63) + (3^3) + \dots,$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$S_k = 1 + (k) + (2k) + (k^2) + (3k) + (2k, k) + (k^3) + \dots,$$

and thence if $U_1, U_2, U_3, \dots$ be the product-sums,

$$U_1 = S_1,$$
$$2!\,U_2 = S_1^2 + S_2,$$
$$3!\,U_3 = S_1^3 + 3S_1S_2 + 2S_3,$$
$$4!\,U_4 = S_1^4 + 6S_1^2S_2 + 3S_2^2 + 8S_1S_3 + 6S_4,$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$m!\,U_m = \Sigma \frac{m!}{m_1!\,m_2!\,m_3!\dots}\left(\frac{S_1}{1}\right)^{m_1}\left(\frac{S_2}{2}\right)^{m_2}\left(\frac{S_3}{3}\right)^{m_3}\dots.$$

This is the expression of the enumerating function U_m in which the coefficient of the function $(pqr\dots)$ is to be taken.

66. If, on development, we find that

$$U_m = \dots + \theta\,(pqr\dots) + \dots,$$

our operating symbols shew us that

$$D_pD_qD_r\dots U_m = \theta D_pD_qD_r\dots(pqr\dots) = \theta,$$

since no other terms on the right-hand side survive the operations. We must therefore learn how to operate with the D symbol. It will be remembered that the symbol D_m causes every symmetric function, whose partition does not involve the number m, to vanish, and that

when the number m does appear it strikes out that number from the partition once. Now the portion of S_1, that involves m in partitions, is

$$(m) + (m1) + (m2) + (m1^2) + (m3) + (m21) + (m1^3) + \dots \text{ ad. inf.}$$

Hence
$$D_m S_1 = S_1,$$

or every operative symbol leaves S_1 unaltered.

Also
$$D_{2m} S_2 = S_2, \quad D_{2m+1} S_2 = 0,$$

since S_2 does not involve any uneven number.

Generally $D_{im} S_i = S_i$ and $D_s S_i = 0$ unless s is a multiple of i.

The effect of D_m upon $S_1^{k_1}$ comes next for consideration. The symbol operates through the compositions of m into k_1 parts, zero being included as a possible part.

Thus for example, omitting the operator D_0 for convenience, replacing $D_0 S_1$ by its value S_1,

$$\begin{aligned} D_3 S_1^3 = {} & D_3 S_1 \,.\, S_1 \,.\, S_1 + S_1 \,.\, D_3 S_1 \,.\, S_1 + S_1 \,.\, S_1 \,.\, D_3 S_1 \\ & + D_2 S_1 \,.\, D_1 S_1 \,.\, S_1 + D_2 S_1 \,.\, S_1 \,.\, D_1 S_1 + S_1 \,.\, D_2 S_1 \,.\, D_1 S_1 \\ & + D_1 S_1 \,.\, D_2 S_1 \,.\, S_1 + D_1 S_1 \,.\, S_1 \,.\, D_2 S_1 + S_1 \,.\, D_1 S_1 \,.\, D_2 S_1 \\ & + D_1 S_1 \,.\, D_1 S_1 \,.\, D_1 S_1 \\ = {} & 10 S_1^3, \end{aligned}$$

and this coefficient of S_1^3 arises because the number of compositions of the number 3 into 3 parts is 10.

This example establishes that the effect of D_m upon $S_1^{k_1}$ is to multiply $S_1^{k_1}$ by a number equal to the number of compositions of m into k_1 parts, zero counting as a part. Hence by Art. 30

$$D_m S_1^{k_1} = \binom{m + k_1 - 1}{m} S_1^{k_1}.$$

67. The effect of D_m upon $S_2^{k_2}$ depends upon the compositions of m into k_2 even parts, zero taking its place as an even part. Hence unless m be even it causes $S_2^{k_2}$ to vanish. Considering then the symbol D_{2m} we observe that the compositions of $2m$ into even parts are equal in number to the whole number of compositions of m, for they are obtainable by multiplying by 2 each part of the latter composition.

Hence
$$D_{2m} S_2^{k_2} = \binom{m + k_2 - 1}{m} S_2^{k_2}.$$

Generally there is no difficulty in establishing that

$$D_{im} S_i^{k_i} = \binom{m + k_i - 1}{m} S_i^{k_i}$$

while the result is zero if the suffix of the operative symbol is not a multiple of i.

68. Finally, consider the value of

$$D_m S_1^{k_1} S_2^{k_2} \, .. \, S_i^{k_i}.$$

In dealing with the compositions of m into $k_1 + k_2 + \dots + k_i$ parts, zero counting as a part and retaining the factors of the operand in the above order, there is no condition that must be fulfilled by the first k_1 parts of the composition; the next k_2 parts must be multiples of 2; the next k_3 parts must be multiples of 3; ..., and finally, the last k_i parts must be multiples of i. Unless these conditions are satisfied the result of the operation derived from the composition will be zero. The complete result of the operation is the multiplication of the operand by an integer equal to the number of the compositions of m which have the properties above set forth. This integer is equal to the coefficient of x^m in the development of the algebraic product

$$(1 + x + x^2 + \dots)^{k_1} (1 + x^2 + x^4 + \dots)^{k_2} \dots (1 + x^i + x^{2i} + \dots)^{k_i},$$

because in the ordered multiplication an exponent of x can only be made up of

k_1	numbers	each	divisible	by	unity
k_2	,,	,,	,,	,,	2
k_3	,,	,,	,,	,,	3
$\vdots$					$\vdots$
k_i	,,	,,	,,	,,	i

Hence $$D_m S_1^{k_1} S_2^{k_2} \dots S_i^{k_i}$$

= coefficient of x^m in

$$(1-x)^{-k_1} (1-x^2)^{-k_2} \dots (1-x^i)^{-k_i} S_1^{k_1} S_2^{k_2} \dots S_i^{k_i}.$$

69. In the light of this result consider the enumeration of the distributions of objects of specification $(pqr \dots)$ into two or fewer similar boxes, or, what is the same question, the enumeration of the partitions of the multipartite number $(pqr \dots)$ into two or fewer parts. By Art. 62 we seek the coefficient of the function $(pqr \dots)$ in the development of

$$U_2 = \frac{1}{2!}(S_1^2 + S_2).$$

This is equal to the *first term in*

$$D_p D_q D_r \dots \frac{1}{2!}(S_1^2 + S_2),$$

which materialises when, after the operations, we put each of the quantities S_1, S_2 equal to unity.

Now by the theorem that has been established in Art. 65

$$D_{p_1}\frac{1}{2\,!}(S_1^2+S_2)$$

is equal to the coefficient of $x_1^{p_1}$ in

$$\frac{1}{2}\left\{\frac{S_1^2}{(1-x_1)^2}+\frac{S_2}{1-x_1^2}\right\}\equiv\frac{1}{2}\frac{(S_1^2+S_2)+x_1(S_1^2-S_2)}{(1-x_1)(1-x_1^2)}.$$

Hence the coefficient of the function (p_1) is, putting $S_1=S_2=1$, the coefficient of $x_1^{p_1}$ in

$$\frac{1}{(1-x_1)(1-x_1^2)}.$$

This number enumerates the partitions of the (unipartite) number p_1 into two or fewer parts and solves the corresponding problem in distributions.

This result is of course well known since the time of Euler.

70. Proceeding from the result

$$D_{p_1}\tfrac{1}{2}(S_1^2+S_2)=\text{coeff. of } x_1^{p_1} \text{ in}$$

$$\frac{1}{2}\left\{\frac{S_1^2}{(1-x_1)^2}+\frac{S_2}{1-x_1^2}\right\}$$

we can further operate with the symbol D_{p_2} and find that

$$D_{p_1}D_{p_2}\tfrac{1}{2}(S_1^2+S_2)=\text{coeff. of } x_1^{p_1}x_2^{p_2} \text{ in}$$

$$\frac{1}{2}\left\{\frac{S_1^2}{(1-x_1)^2(1-x_2)^2}+\frac{S_2}{(1-x_1^2)(1-x_2^2)}\right\},$$

shewing us that the coeff. of the function (p_1p_2) is equal to the coeff. of $x_1^{p_1}x_2^{p_2}$ in

$$\frac{1}{2}\left\{\frac{1}{(1-x_1)^2(1-x_2)^2}+\frac{1}{(1-x_1^2)(1-x_2^2)}\right\}$$

$$\equiv\frac{1+x_1x_2}{(1-x_1)(1-x_1^2).(1-x_2)(1-x_2^2)}.$$

This number enumerates the partitions of the bipartite number (p_1p_2) into two or fewer parts and solves the corresponding problem in distributions.

Further, if we denote by $P(pq, 2)$, $P(p, 2)$ the numbers of the partitions of (pq) and (p) into two or fewer parts we see that we may write

$$P(p_1p_2, 2) = P(p_1, 2)\,P(p_2, 2) + P(p_1 - 1, 2)\,P(p_2 - 1, 2),$$

a convenient formula. As an example

$$P(33, 2) = \{P(3, 2)\}^2 + \{P(2, 2)\}^2,$$

and observing that the numbers 3, 2 have each of them 2 partitions into 2 or fewer parts

$$P(33, 2) = 2^2 + 2^2 = 8.$$

The 8 partitions, thus enumerated, are

$$\begin{array}{cccc} (33), & (32, 01), & (23, 10), & (31, 02), \\ (13, 20), & (22, 11), & (21, 12), & (30, 03). \end{array}$$

In general, since $P(2p, 2) = p + 1 = P(2p+1, 2)$, we have the formulae

$$P(2p_1, 2p_2, 2) = (p_1 + 1)(p_2 + 1) + p_1p_2,$$
$$P(2p_1, 2p_2 + 1, 2) = (2p_1 + 1)(p_2 + 1),$$
$$P(2p_1 + 1, 2p_2 + 1, 2) = 2(p_1 + 1)(p_2 + 1).$$

71. For the multipartite number $(p_1p_2 \ldots p_s)$ we find that

$D_{p_1}D_{p_2} \ldots D_{p_s}\frac{1}{2}(S_1^2 + S_2) = \text{coeff. of } x_1^{p_1}x_2^{p_2} \ldots x_s^{p_s}$ in

$$\frac{1}{2}\left\{\frac{S_1^2}{(1-x_1)^2(1-x_2)^2\ldots(1-x_s)^2} + \frac{S_2}{(1-x_1^2)(1-x_2^2)\ldots(1-x_s^2)}\right\},$$

and thence we establish that the partitions of the multipartite number $(p_1p_2 \ldots p_s)$, into two or fewer parts, are enumerated by the coeff. of $x_1^{p_1}x_2^{p_2} \ldots x_s^{p_s}$ in

$$\frac{1}{2}\left\{\frac{1}{(1-x_1)^2(1-x_2)^2\ldots(1-x_s)^2} + \frac{1}{(1-x_1^2)(1-x_2^2)\ldots(1-x_s^2)}\right\},$$

or in
$$\frac{1 + \Sigma x_1x_2 + \Sigma x_1x_2x_3x_4 + \ldots}{(1-x_1)(1-x_1^2)\,.\,(1-x_2)(1-x_2^2)\ldots(1-x_s)(1-x_s^2)},$$

the last numerator term being $\Sigma x_1x_2 \ldots x_{s-1}$ or $\Sigma x_1x_2 \ldots x_s$, according as s is uneven or even.

From this result general formulae may be constructed as above for the particular case $s = 2$.

72. Passing to the partitions into three or fewer parts we have

$$U_3 - \tfrac{1}{6}(S_1^3 + 3S_1S_2 + 2S_3),$$

$D_{p_1}U_3$ equal to the coeff. of $x_1^{p_1}$ in

$$\frac{1}{6}\left\{\frac{S_1^3}{(1-x_1)^3} + 3\frac{S_1S_2}{(1-x_1)(1-x_1^2)} + 2\frac{S_3}{1-x_1^3}\right\},$$

and thence the coeff. of the function (p_1) in U_3 is equal to the coeff. of $x_1^{p_1}$ in

$$\frac{1}{6}\left\{\frac{1}{(1-x_1)^3} + 3\frac{1}{(1-x_1)(1-x_1^2)} + 2\frac{1}{1-x_1^3}\right\},$$

or in

$$\frac{1}{(1-x_1)(1-x_1^2)(1-x_1^3)},$$

the well-known result in the case of the unipartite numbers.

Similarly for the partitions of bipartite numbers we are led to the coeff. of $x_1^{p_1}x_2^{p_2}$ in

$$\frac{1}{6}\left\{\frac{1}{(1-x_1)^3.(1-x_2)^3} + 3\frac{1}{(1-x_1)(1-x_1^2).(1-x_2)(1-x_2^2)} + 2\frac{1}{(1-x_1^3).(1-x_2^3)}\right\},$$

or in

$$\frac{1 + x_1x_2 + x_1^2x_2 + x_1x_2^2 + x_1^2x_2^2 + x_1^3x_2^3}{(1-x_1)(1-x_1^2)(1-x_1^3).(1-x_2)(1-x_2^2)(1-x_2^3)}$$

73. In general for the case of the partitions of s-partite numbers into three or fewer parts we are led to the coeff. of $x_1^{p_1}x_2^{p_2}\dots x_s^{p_s}$ in

$$\frac{1}{6}\left\{\frac{1}{(1-x_1)^3(1-x_2)^3\dots(1-x_s)^3} + 3\frac{1}{(1-x_1)(1-x_1^2).(1-x_2)(1-x_2^2)\dots(1-x_s)(1-x_s^2)} + 2\frac{1}{(1-x_1^3)(1-x_2^3)\dots(1-x_s^3)}\right\}.$$

In a similar manner the enumerating generating function for the partitions of the multipartite numbers $(p_1p_2\dots p_s)$ into m or fewer parts can be constructed and the general problem of distribution before us may be regarded as solved.

74. For the partitions of the unipartite number p_1 into m or fewer parts the generating function comes out, after simplification, in the Eulerian form

$$\frac{1}{(1-x_1)(1-x_1^2)\dots(1-x_1^m)}.$$

75. The final and most general case of distribution presents itself when the objects have the specification $(p_1 p_2 \dots p_s)$ and the boxes the specification $(m_1 m_2 \dots m_t)$, the whole number of the boxes being any number not greater than the whole number of the objects.

Here the enumerating generating function is

$$U_{m_1} U_{m_2} \dots U_{m_t},$$

in which we seek the coeff. of the function

$$(p_1 p_2 \dots p_s).$$

For consider any distribution of the objects into the boxes. It consists of objects having a certain specification distributed into boxes of specification (m_1), together with objects of other specifications distributed into boxes of specifications (m_2), (m_3), ... (m_t) respectively. The aggregate of these specifications of combinations of objects constitutes a composition of the multipartite numbers $(p_1 p_2 \dots p_s)$ into t or fewer parts, multipartite parts consisting wholly of zeros being admissible. Since any combination of objects may appertain to any set and the sets are not interchangeable, we obtain the generating function by simply multiplying together the generating functions which belong to the separate sets of boxes.

76. The application of this theorem to the distribution of objects of specification (p_1) is interesting. The enumerating function is

$$\begin{aligned} &\frac{1}{(1-x)(1-x^2)\dots(1-x^{m_1})}\\ &\times\frac{1}{(1-x)(1-x^2)\dots(1-x^{m_2})}\\ &\times\frac{1}{(1-x)(1-x^2)\dots(1-x^{m_3})}\\ &\times\dots\dots\dots\dots\dots\dots\\ &\times\frac{1}{(1-x)(1-x^2)\dots(1-x^{m_t})}\\ &\equiv\frac{1}{(1-x)^{n_1}(1-x^2)^{n_2}(1-x^3)^{n_3}\dots},\end{aligned}$$

where the succession of numbers $n_1, n_2, n_3, \dots$ is related to the succession $m_1, m_2, m_3, \dots m_t$ in the following manner.

We write down $m_1, m_2, m_3, \dots m_t$ units in succession

$$\begin{array}{llll} 1\ 1\ 1\ 1 \dots m_1 & \text{units} \\ 1\ 1\ 1\ 1 \dots m_2 & ,, \\ 1\ 1\ 1\ 1 \dots m_3 & ,, \\ \dots\dots\dots\dots & \\ 1\ 1\ 1\ 1 \dots m_t & ,, \end{array}$$

the numbers $m_1, m_2, m_3 \ldots$ being assumed in descending order of magnitude and then add by columns producing a partition $(n_1, n_2, n_3, \ldots)$ which is said to be conjugate to $(m_1 m_2 \ldots m_t)$.

We have therefore a remarkable theorem:—

"The number of distributions of objects of specification (p) into boxes of specification $(m_1 m_2 \ldots m_t)$ is given by the coeff. of x^p in the function

$$(1-x)^{-n_1}(1-x^2)^{-n_2}(1-x^3)^{-n_3}\ldots$$

where $(n_1 n_2 n_3 \ldots)$ is the partition conjugate to $(m_1 m_2 \ldots m_t)$."

77. As a verification observe that if $(m_1 m_2 \ldots m_t) \equiv (m)$ the conjugate partition is (1^m) and the enumerating function is

$$(1-x)^{-1}(1-x^2)^{-1}\ldots(1-x^m)^{-1},$$

whereas if $(m_1 m_2 \ldots m_t) \equiv (1^m)$ the conjugate partition is (m) and the enumerating function is

$$(1-x)^{-m}.$$

As another example suppose $(m_1 m_2 \ldots m_t) \equiv (221)$; the conjugate partition is (32) and the enumerating function is

$$(1-x)^{-3}(1-x^2)^{-2},$$

which is $$1 + 3x + 8x^2 + 16x^3 + 30x^4 + \ldots.$$

The distributions of the assemblages α^2, α^3 are

A	A	B	B	C
α^2	•	•	•	•
•	•	α^2	•	•
•	•	•	•	α^2
α	α	•	•	•
α	•	α	•	•
α	•	•	•	α
•	•	α	α	•
•	•	α	•	α

No. = 8

A	A	B	B	C
α^3	•	•	•	•
•	•	α^3	•	•
•	•	•	•	α^3
α^2	α	•	•	•
α^2	•	α	•	•
α^2	•	•	•	α
•	•	α^2	α	•
•	•	α^2	•	α
α	•	α^2	•	•
α	•	•	•	α^2
•	•	α	•	α^2
α	α	α	•	•
α	α	•	•	α
α	•	α	α	•
α	•	α	•	α
•	•	α	α	α

No. = 16

78. If we restrict the symmetric functions utilised so that no part greater than k appears the effect is to restrict the distributions to the extent that not more than k similar objects can appear in any one box.

We may usefully examine the case $k = 1$.

Instead of the functions $S_1, S_2, \dots$ we take

$$A_1 = 1 + (1) + (1^2) + (1^3) + \dots,$$
$$A_2 = 1 + (2) + (2^2) + (2^3) + \dots,$$
$$\dots\dots\dots\dots\dots\dots$$
$$A_m = 1 + (m) + (m^2) + (m^3) + \dots;$$

and then $U_1 = A_1$,

$$2!\, U_2 = A_1^2 + A_2,$$
$$3!\, U_3 = A_1^3 + 3A_1A_2 + 2A_3,$$
$$\dots\dots\dots\dots\dots\dots$$
$$m!\, U_m = \Sigma \frac{m!}{m_1!\, m_2!\, m_3! \dots} \left(\frac{A_1}{1}\right)^{m_1} \left(\frac{A_2}{2}\right)^{m_2} \left(\frac{A_3}{3}\right)^{m_3} \dots.$$

79. For the operation of the D symbol we have

$$D_1 A_1 = A_1,\ D_0 A_1 = A_1,\ D_m A_1 = 0 \text{ in other cases},$$

and generally $\quad D_s A_m = A_m$ when $s = m$ or zero,

$$D_s A_m = 0 \text{ in every other case.}$$

Also the symbol, operating through the composition of its suffix into units, yields

$$D_k A_1^m = \binom{m}{k} A_1^m;\quad D_{sk} A_s^m = \binom{m}{k} A_s^m.$$

For the operand

$$A_1^{m_1} A_2^{m_2} \dots A_i^{m_i}$$

the symbol D_s operates through the compositions of s into

$$m_1 + m_2 + \dots + m_i \text{ parts},$$

zero counting as a part. From the law of operation given above it can be seen that for the operation associated with such a composition to have an effect other than zero,

the first m_1	parts must be	zero or	unity,
next m_2	,,	,,	two
,, m_3	,,	,,	three
⋮ ⋮	,,	,,	⋮
,, m_i	,,	,,	i

The number of such compositions is the coefficient of x^s in the product

$$(1+x)^{m_1}(1+x^2)^{m_2}(1+x^3)^{m_3}\dots(1+x^i)^{m_i}$$

as is evident when the orderly multiplication is carried out (cf. Art. 14). Thence

$$D_s A_1^{m_1} A_2^{m_2}\dots A_i^{m_i}$$
$$= A_1^{m_1} A_2^{m_2}\dots A_i^{m_i} \times \text{coefficient of } x^s \text{ in } (1+x)^{m_1}(1+x^2)^{m_2}\dots(1+x^i)^{m_i}\,*.$$

80. To apply this result, consider the distributing of objects of specification $(2^{k_2}1^{k_1})$ into two or fewer similar boxes—in other words, the partitions of the multipartite number $(2^{k_2}1^{k_1})$ into two or fewer parts subject to the restriction that no box is to contain two similar objects—or no constituent of the multipartite parts to involve numbers greater than unity.

We find that

$$D_2U_2 = D_2(\tfrac{1}{2}A_1^2 + \tfrac{1}{2}A_2) = \tfrac{1}{2}A_1^2 + \tfrac{1}{2}A_2,$$

because the coefficients of x^2 in $(1+x)^2$ and in $(1+x^2)$ are both unity. Hence

$$D_2^{k_2}U_2 = \tfrac{1}{2}A_1^2 + \tfrac{1}{2}A_2.$$

Now $$D_1U_2 = D_1(\tfrac{1}{2}A_1^2 + \tfrac{1}{2}A_2) = A_1^2,$$

because the coefficients of x in $(1+x)^2$ and in $(1+x^2)$ are 2 and zero respectively.

Hence by repeated operation

$$D_2^{k_2}D_1^{k_1}U_2 = 2^{k_1-1}A_1^2,$$

establishing that the coefficient of the function $(2^{k_2}1^{k_1})$ in U_2 is

$$2^{k_1-1}.$$

Ex. gr. Suppose that the objects for distribution are

$$\alpha\ \alpha\ \beta\ \beta\ \gamma\ \gamma\ \delta\ \epsilon\ \theta,$$

so that $$k_2 = k_1 = 3.$$

The distributions—four in number—are

A	A
$\alpha\beta\gamma\delta\epsilon\theta$	$\alpha\beta\gamma$
$\alpha\beta\gamma\delta\epsilon$	$\alpha\beta\gamma\theta$
$\alpha\beta\gamma\delta\theta$	$\alpha\beta\gamma\epsilon$
$\alpha\beta\gamma\epsilon\theta$	$\alpha\beta\gamma\delta$

* The reader will observe that when the magnitude of the parts of the partitions is restricted to be not greater than the integer k, the corresponding function of x is

$$(1+x+x^2+\dots+x^k)^{m_1}(1+x^2+x^4+\dots+x^{2k})^{m_2}\dots(1+x^i+x^{2i}+\dots+x^{ki})^{m_i}$$
$$\equiv\left(\frac{1-x^{k+1}}{1-x}\right)^{m_1}\left(\frac{1-x^{2k+2}}{1-x^2}\right)^{m_2}\dots\left(\frac{1-x^{ik+i}}{1-x^i}\right)^{m_i}$$

Again, let the objects be of specification $(3^{k_3}2^{k_2}1^{k_1})$ and let there be three or fewer similar boxes, the distributions being subject to the same restriction as before.

We find that

$$D_3 \tfrac{1}{6}(A_1^3 + 3A_1A_2 + 2A_3) = \tfrac{1}{6}(A_1^3 + 3A_1A_2 + 2A_3),$$

because the coefficients of x^3 in

$$(1+x)^3,\ (1+x)(1+x^2),\ (1+x^3)$$

are all equal to unity.

Hence $\qquad D_3^{k_3}$ gives $\tfrac{1}{6}(A_1^3 + 3A_1A_2 + 2A_3)$.

Now the coefficients of x^2 in the three functions of x are 3, 1 and 0.

Hence $\qquad D_2 \tfrac{1}{6}(A_1^3 + 3A_1A_2 + 2A_3) = \tfrac{1}{2}(A_1^3 + A_1A_2)$

and $\qquad D_3^{k_3}D_2^{k_2}\tfrac{1}{6}(A_1^3 + 3A_1A_2 + 2A_3) = \tfrac{1}{2}(3^{k_2-1}A_1^3 + A_1A_2)$,

and finally

$$D_3^{k_3}D_2^{k_2}D_1^{k_1}\tfrac{1}{6}(A_1^3 + 3A_1A_2 + 2A_3) = \tfrac{1}{2}(3^{k_2+k_1-1}A_1^3 + A_1A_2),$$

establishing that the coefficient of the function $(3^{k_3}2^{k_2}1^{k_1})$ in U_3 is

$$\tfrac{1}{2}(3^{k_2+k_1-1}+1).$$

Ex. gr. If the objects to be distributed are

$$\alpha\alpha\alpha\ \ \beta\beta\beta\ \ \gamma\gamma\ \ \delta\delta\ \ \epsilon\epsilon\ \ \theta,$$

so that $\qquad k_3 = 2,\ \ k_2 = 3,\ \ k_1 = 1$

we have the fourteen distributions

A	A	A	A	A	A
$\alpha\beta\gamma\delta\epsilon\theta$	$\alpha\beta\gamma\delta\epsilon$	$\alpha\beta$	$\alpha\beta\gamma\delta\epsilon$	$\alpha\beta\gamma\delta\epsilon$	$\alpha\beta\theta$
$\alpha\beta\gamma\delta\epsilon\theta$	$\alpha\beta\delta\epsilon$	$\alpha\beta\gamma$	$\alpha\beta\gamma\epsilon\theta$	$\alpha\beta\delta\epsilon$	$\alpha\beta\gamma\delta$
$\alpha\beta\gamma\delta\epsilon\theta$	$\alpha\beta\gamma\epsilon$	$\alpha\beta\delta$	$\alpha\beta\gamma\delta\epsilon$	$\alpha\beta\delta\theta$	$\alpha\beta\gamma\epsilon$
$\alpha\beta\gamma\delta\epsilon\theta$	$\alpha\beta\gamma\delta$	$\alpha\beta\epsilon$	$\alpha\beta\gamma\delta\epsilon$	$\alpha\beta\delta\epsilon$	$\alpha\beta\gamma\theta$
$\alpha\beta\gamma\delta\epsilon$	$\alpha\beta\delta\epsilon\theta$	$\alpha\beta\gamma$	$\alpha\beta\delta\epsilon\theta$	$\alpha\beta\gamma\delta$	$\alpha\beta\gamma\epsilon$
$\alpha\beta\gamma\epsilon\theta$	$\alpha\beta\gamma\delta\epsilon$	$\alpha\beta\delta$	$\alpha\beta\gamma\delta\theta$	$\alpha\beta\gamma\epsilon$	$\alpha\beta\delta\epsilon$
$\alpha\beta\gamma\delta\epsilon$	$\alpha\beta\gamma\delta\theta$	$\alpha\beta\epsilon$	$\alpha\beta\gamma\delta\epsilon$	$\alpha\beta\gamma\delta$	$\alpha\beta\epsilon\theta$

81. In general to shew the nature of the theorems more clearly we observe that

$$D_{p_1}D_{p_2}\dots D_{p_s}U_2 = \text{coefficient of } x_1^{p_1}x_2^{p_2}\dots x_s^{p_s} \text{ in}$$
$$\tfrac{1}{2}\{(1+x_1)(1+x_2)\dots(1+x_s)\}^2A_1^2$$
$$+\tfrac{1}{2}(1+x_1^2)(1+x_2^2)\dots(1+x_s^2)A_2,$$

so that the enumerating function is

$$\tfrac{1}{2}\{(1+x_1)(1+x_2)\dots(1+x_s)\}^2$$
$$+\tfrac{1}{2}(1+x_1^2)(1+x_2^2)\dots(1+x_s^2),$$

and similarly, derived from

$$D_{p_1}D_{p_2}\dots D_{p_s}U_3,$$

we obtain the enumerating function

$$\tfrac{1}{6}\{(1+x_1)(1+x_2)\dots(1+x_s)\}^3$$
$$+\tfrac{1}{2}\{(1+x_1)(1+x_2)\dots(1+x_s)$$
$$\times(1+x_1^2)(1+x_2^2)\dots(1+x_s^2)\}$$
$$+\tfrac{1}{3}(1+x_1^3)(1+x_2^3)\dots(1+x_s^3),$$

and so on in the higher cases.

In conclusion it will be clear that an important part of the Theory of Combinations and Permutations is intimately connected with the Theory of Symmetric Functions in elementary algebra.

In Combinatory Analysis, by the author, the correspondence is carried much further and it is shewn that either theory is a powerful instrument of research in the other. The fact is that in theorems of Combinations and Permutations the entities dealt with come into consideration in a symmetrical manner and a symmetrical method of investigation is at once suggested. Moreover it will be found in nearly every case that the appropriate method, though it may be in appearance devoid of symmetry, is when sufficiently examined, symmetrical. The binomial coefficients which enter largely into combinatory theorems are themselves symmetric functions of zero weight. Ex. gr.

$$\Sigma\alpha^0 \quad = (0) = \binom{n}{1}$$
$$\Sigma\alpha^0\beta^0 \quad = (0^2) = \binom{n}{2}$$
$$\Sigma\alpha^0\beta^0\gamma^0 = (0^3) = \binom{n}{3}$$
$$\vdots$$

the number of the quantities α, β, γ, ... being n.

There is an algebra of these numerical functions which deals with their representation by means of partitions with zero parts and we find an appropriate operator

$$D_0$$

which operates through the compositions of zero into zero parts. These

are identical with the partitions of zero into zero parts and are infinite in number, viz. :—

0, 00, 000, 0000, ... ad inf.

It has been noted that the operator

$$D_m$$

where m is a positive integer > 0 is in fact a partial differential operator of the order m. When $m = 0$, we find that the operator is one that is met with in the Calculus of Finite Differences.

There is throughout a corresponding theory of the enumeration of numbered diagrams of the 'Magic Square' type which has been much developed and will without doubt be the subject of further investigations.

THREE LECTURES ON FERMAT'S LAST THEOREM

BY

L. J. MORDELL

MANCHESTER COLLEGE OF TECHNOLOGY

PREFACE

IN March 1920, I gave at Birkbeck College, London, a course of three public lectures on Fermat's Last Theorem. The lectures were intended primarily for persons with a mathematical training, but not necessarily for those who had made a special study of the Theory of Numbers. A general account was given of the various methods that have been devised for dealing with the question, more attention being paid to principles than to details.

This booklet consists of the lectures in practically the form in which they were delivered. It also includes a few details which it was found convenient to omit from the lectures. I hope it may be of assistance in giving to the reader some idea, not only of the difficulties involved, but also of the progress made in dealing with this famous theorem.

I have to acknowledge my indebtedness not only to the authors mentioned herein, but also to the works of Smith, Bachmann, Hilbert, Kronecker, Sommer, and Dickson, on the Theory of Numbers. Full references to the subject are given by Dickson in his very useful paper on "Fermat's Last Theorem" in the *Annals of Mathematics*, Vol. XVIII. 1917; and in Vol. II. of his *History of the Theory of Numbers*, which has just been published.

L. J. MORDELL.

November 1920.

CONTENTS

CHAPTER I

STATEMENT OF THE THEOREM

Of all the abstract sciences, perhaps none is so remarkable for the ease with which theorems are arrived at inductively, and for the difficulty and importance of the developments arising in the efforts to prove the theorems so suggested, as the Theory of Numbers. An admirable illustration of this fact is furnished by Fermat's Last Theorem, namely, that if n is a positive integer greater than two, the equation

$$x^n + y^n = z^n \quad \text{..............................} (1)$$

cannot be satisfied by integer values for the unknowns x, y, and z unless one of them is zero. On the contrary, when $n = 2$, it is well known that the equation possesses an infinite number of solutions in integers.

Fermat (1601—1665) was a French mathematician of the first rank, who made a special study of the Theory of Numbers, including that part of the subject dealing with the solution of indeterminate equations, called Diophantine Analysis, after the Greek mathematician Diophantus who flourished during* the third century A.D. A new edition of the latter's works was brought out by Bachet in 1621. Fermat possessed a copy of this work and entered in the margin of the pages a number of theorems he had discovered, most of which are now included as special cases in the classical theory of the subject, but without any proofs or indications of his methods. Besides the theorem now known as his last theorem, he placed the remark that he had discovered a truly wonderful proof but that the margin of the book was too small to contain it. Since that time, no general proof has been found for all values of n, even though it has been attempted by the greatest of mathematicians including Euler, Legendre, Gauss, Abel, Dirichlet, Cauchy and Kummer, has been several times made the prize question of learned societies such as the Academies at Paris and Brussels, and though finally, in 1907, a prize of 100,000 marks was established for the first proof.

* *Diophantus of Alexandria*, by Sir Thomas L. Heath, 2nd edition, p. 2.

Did Fermat prove his Theorem ?

The question immediately suggests itself.—Is it probable, that nearly three centuries ago, Fermat really proved this theorem, which still baffles mathematicians who have at their disposal the wonderful and far reaching developments in mathematics since Fermat's time—especially as it seems likely that Fermat's methods could only be elementary considered from a modern standpoint ? From what is known of Fermat's character, it is fairly certain that at any rate he was under the impression that he had a proof meriting his description of it. This statement is confirmed by the fact that when enunciating a theorem to the effect that $2^{2^n}+1$ was a prime number for all positive integral values of n, he added that while convinced of the truth of this theorem, he could not prove it*. Many years afterwards Euler showed that the theorem was false, and that 641 was a factor when $n=5$.

It is of course possible that Fermat was mistaken in thinking that his proof was valid, for even the greatest of mathematicians have made mistakes. The late Prof. H. J. S. Smith, while pointing out that Gauss was unfavourably inclined to Fermat, thought however that there was no ground for supposing that Fermat was mistaken.

Analysis of another statement by Fermat

A little light perhaps may be thrown on Fermat's statement by considering a similar case. He had proposed as a problem to the English mathematicians to show that there was only one integer solution of the equation

$$y^3 = x^2 + 2,$$

obviously $x = \pm 5$, $y = 3$. On this he has a note to the effect that there was no difficulty in finding a solution in fractions, but that he had discovered an entirely new method, wonderfully beautiful and most subtle, which enabled him to solve such questions in integers. This statement seems clear and straightforward and would lead one to suppose that, given an equation of the form

$$y^3 = x^2 + k,$$

where k is an integer, Fermat possessed a method which enabled him to ascertain if the equation possessed integer solutions, and in that case to find them.

* See however Dickson's *History of the Theory of Numbers*, Vol. I. p. 375.

Fermat's statement, however, cannot be as clear as it seems to be. For I showed* several years ago that no equation of the form

$$y^3 = x^2 + k,$$

where k is a positive or negative integer, could have more than a finite number of integer solutions. It seems unlikely that this fact was known to Fermat, so we are led to conjecture that his method must have been equivalent to some such process as the following.

In $$y^3 = x^2 + 2,$$

put $$y = a^2 + 2b^2,$$

where a and b are integers, and take

$$x + \sqrt{-2} = (a + b\sqrt{-2})^3.$$

Equating real and imaginary parts

$$x = a^3 - 6ab^2,$$

$$1 = b(3a^2 - 2b^2).$$

Since a and b are integers

$$b = \pm 1, \quad 3a^2 - 2b^2 = \pm 1,$$

or $$b = \pm 1, \quad a = \pm 1,$$

giving $$x = \pm 5, \quad y = 3.$$

It is of course by no means obvious that this process, which can be described without the use of complex quantities, gives all the values of x and y. In any case it seems doubtful if Fermat's description of his method would be justified at the present time.

At the same time it is possible that Fermat did possess a valid proof of his last theorem, but from the circumstances of the case, it is extremely difficult for us to form any conception of his method. One can easily recall a number of theorems which have proved extraordinarily difficult to great mathematicians and which now seem elementary enough. Nothing can appear simpler than the solution of the cubic equation, but many centuries elapsed between the solution of the quadratic and of the cubic. Another instance of a different type is supplied by the proof of the transcendental character of π, i.e. the impossibility of solving the problem commonly called the

* A statement by Fermat, *Proceedings of the London Mathematical Society* (Read Feb. 1918), (Records, etc.), Ser. 2, Vol. XVIII. (1919), pp. v, vi. The same result holds for the equation

$$ey^2 = ax^3 + bx^2 + cx + d,$$

where a, b, c, d, e are any integers for which the right-hand side has no squared factors in x.

squaring of the circle, which can now be put in a very simple and elementary form.

Mathematical study and research are very suggestive of mountaineering. Whymper made seven efforts before he climbed the Matterhorn in the 1860's and even then it cost the lives of four of his party. Now, however, any tourist can be hauled up for a small cost, and perhaps does not appreciate the difficulty of the original ascent. So in mathematics, it may be found hard to realise the great initial difficulty of making a little step which now seems so natural and obvious, and it may not be surprising if such a step has been found and lost again.

A simplification of the Problem

Coming back to the equation

$$x^n + y^n = z^n \quad \ldots\ldots\ldots\ldots\ldots\ldots(1),$$

it is obvious that if any two of the unknowns have a common factor k, then the third unknown is also divisible by k. Putting

$$x,\ y,\ z = k\xi,\ k\eta,\ k\zeta,$$

respectively, in equation (1), k^n divides out, leaving

$$\xi^n + \eta^n = \zeta^n,$$

where now no two of the unknowns have a common factor. There will be no loss of generality then if it is supposed that no two of the unknowns in the original equation (1) have a common factor. Next it is sufficient to prove the impossibility of equation (1) when n is equal to 4 or to any odd prime p. For if each of the equations

$$x^4 + y^4 = z^4, \quad x^p + y^p = z^p \quad \ldots\ldots\ldots\ldots(2)$$

is insoluble, the same holds of equation (1) since n must be divisible by either 4 or an odd prime. For example if n is divisible by p, say $n = pq$, equation (1) can be written as

$$(x^q)^p + (y^q)^p = (z^q)^p,$$

which is then insoluble because of the special case (2).

The equation $x^2 + y^2 = z^2$

As regards the case $n = 2$, it is well known that the general solution of

$$x^2 + y^2 = z^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(3),$$

wherein no two of the unknowns have a common factor, is given by

$$x = a^2 - b^2, \quad y = 2ab, \quad z = a^2 + b^2,$$

where y is that one of the unknowns which is even, and a and b are prime to each other and not both odd. This result was known to the Indian mathematicians.

THE EQUATION $x^4 + y^4 = z^4$

The case $n = 4$ is remarkable not only from the fact that in contradistinction to all the other values of n, the theorem can be rigorously proved by absolutely elementary means, that is by methods which do not implicitly make use of new ideas unknown during Fermat's time, but also from the fact that a proof by Fermat for a very closely related theorem is extant. A proof was given by Leibnitz in a manuscript dated December 1678, and also by Euler.

The proof of the theorem is so simple that it will be worth while giving it completely.

It is obviously sufficient to consider the equation

$$x^4 + y^4 = z^2,$$

where x, y and z are all prime to each other. Further it may be assumed that all the quantities referred to are positive. As all numbers are either odd or even, x is of the form $2m$ or $2m + 1$, where m is an integer. Hence x^2 is of the form $4m^2$ or $4m^2 + 4m + 1$, that is of the form $4M$ or $4M + 1$, so that a number of the form $4M + 2$ or $4M + 3$ cannot be a square. Hence x and y cannot both be odd, for then the sum of their fourth powers would be of the form $4M + 2$, and this cannot be a square. Hence either x or y must be even, and as it is obviously immaterial which one is, suppose it is y. Since

$$(x^2)^2 + (y^2)^2 = z^2,$$

it follows from equation (3) that we must have

$$x^2 = a^2 - b^2, \quad y^2 = 2ab, \quad z = a^2 + b^2,$$

where a and b are prime to each other, and not both odd. From

$$x^2 = a^2 - b^2$$

we see that a cannot be even, for then b would be odd and x^2 would be of the form $4M + 3$, which is impossible. We have then

$$x^2 + b^2 = a^2,$$

where b is even, a is odd and prime to b, so that no two of a, b, x have a common factor. Hence it follows from equation (3) that

$$x = p^2 - q^2, \quad b = 2pq, \quad a = p^2 + q^2,$$

where p and q are prime to each other and not both odd. From

$$y^2 = 2ab,$$

we have

$$y^2 = 4pq\,(p^2 + q^2).$$

Since p and q are prime to each other, each of them is prime to p^2+q^2, and hence all three must be perfect squares. Put then

$$p=r^2, \quad q=s^2, \quad p^2+q^2=t^2,$$

from which

$$r^4+s^4=t^2.$$

Now the values of x, y, z in terms of r, s, t are given by

$$x=r^4-s^4, \quad y=2rst, \quad z=a^2+b^2=r^8+6r^4s^4+s^8,$$

so that

$$z>(r^4+s^4)^2>t^4 \quad \text{or} \quad t<\sqrt[4]{z}.$$

It follows then that if one solution of the equation

$$x^4+y^4=z^2$$

is known for which none of the unknowns is zero, another solution (r, s, t) can be found for which none of the unknowns is zero and such that $t<\sqrt[4]{z}$. This process can be continued, so that an infinite number of positive integers t, t_1, t_2 ... can be found such that

$$t_1<\sqrt[4]{t}, \quad t_2<\sqrt[4]{t_1} \ldots,$$

which is clearly absurd.

This proves the impossibility in the case of $n=4$, the method of proof being known as the method of infinite descent.

THE EQUATION $x^3+y^3=z^3$

The case $n=3$, that is the equation

$$x^3+y^3=z^3 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4),$$

had been known to the Arabian mathematicians nearly seven hundred years before the time of Fermat, and a faulty proof of the impossibility had beeu given by them. It is very probable that Fermat discovered this special case before he discovered the general theorem, for he had proposed as a problem "to find values of x, y, and z satisfying the equation," and had later declared it was impossible. Euler was the first to prove the theorem for this special case, but his proof was incomplete in respect of an assumption wherein lay the real difficulty of the question, and which contained the germ of the development of the theory of ideals which was to be applied so successfully by Kummer many years later. Euler's proof as given in his Algebra is substantially as follows.

Two of the unknowns x, y, z must be odd, and as any of the unknowns may be either positive or negative, there is no loss of generality in supposing that z is even, and that x and y are both odd.

Write then $$x + y = 2p, \quad x - y = 2q$$
so that $$x = p + q, \quad y = p - q,$$
and the original equation becomes
$$2p\,(p^2 + 3q^2) = z^3.$$
Now p and q are prime to each other, and cannot both be odd, for then x and y would not be prime to each other. Further p cannot be odd and q even, for then z^3 would be divisible by 2 and not by 8, which is impossible. Hence p must be even and q odd, so that $p^2 + 3q^2$ is odd. Hence as p and q are prime to each other,
$$2p \text{ and } p^2 + 3q^2$$
are either prime to each other, or have a common factor 3. In the first case p and hence z are both prime to 3, while in the latter case they are both divisible by 3.

Let us consider the first case in detail. As $2p$ and $p^2 + 3q^2$ are prime to each other, each must be a perfect cube, so that we can write
$$p^2 + 3q^2 = r^3 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5).$$
Values of p, q, r can be found by taking
$$r = m^2 + 3n^2,$$
where m and n are integers, and writing
$$p + q\sqrt{-3} = (m + n\sqrt{-3})^3.$$
By equating real and imaginary parts
$$p = m^3 - 9mn^2, \quad q = 3m^2n - 3n^3,$$
and if m and n are prime to each other and not both odd, and m is not divisible by 3, then p and q are prime to each other and p is not divisible by 3. But though this method gives suitable values of p, q, r satisfying
$$p^2 + 3q^2 = r^3,$$
it is by no means obvious that all the values of p, q, r can be found in this way, though as a matter of fact it is so in this particular case. If the equation had been
$$p^2 + 11q^2 = r^3,$$
all the values of p and q would not be given by putting
$$p + q\sqrt{-11} = (m + n\sqrt{-11})^3.$$
The removal of the difficulty involves the study of the arithmetical theory of the binary quadratic form, or of ideal numbers.

Now since $2p$ is a cube, the values of m and n are such that

$$2m(m+3n)(m-3n)$$

is a perfect cube.

But since $$q = 3n(m+n)(m-n)$$

is odd, n is odd and m is even. Hence since m is prime to 3, no two of $2m$, $m+3n$, $m-3n$ can have a common factor; and since their product is a perfect cube, each of them must be a cube. Put then

$$m + 3n = a^3, \quad m - 3n = b^3, \quad 2m = c^3,$$

so that by addition $$a^3 + b^3 = c^3.$$

Hence $$z^3 = 2p(p^2 + 3q^2) = a^3b^3c^3(m^2 + 3n^2)^3,$$

or $$z = abc(m^2 + 3n^2) = \tfrac{1}{4}abc(a^6 + a^3b^3 + b^6),$$

so that as a and b cannot both be unity, z is numerically greater than c. It follows then, just as in the case when $n = 4$, that we should have an infinite sequence of numerically decreasing integers, which is impossible.

The same result follows in the second case when z is divisible by 3, but we need not go into details.

THE EQUATIONS $x^5 + y^5 = z^5$ AND $x^7 + y^7 = z^7$

The next cases to be proved were when $n = 5$ or 7. The first case was dealt with by Legendre and Dirichlet in 1825, while the case of $n = 7$ was proved in 1840 by Lamé and Lebesgue. The proofs involved ideas not greatly dissimilar from the case when $n = 3$ and depended upon two facts. Firstly, that if p is a prime and x and y are prime to each other, the two expressions

$$x + y \text{ and } \frac{x^p + y^p}{x + y}$$

are either prime to each other, or have as a common factor the first power only of p. The proof is immediate, for putting

$$x + y = s,$$

the two expressions become

$$s \text{ and } \frac{x^p + (s-x)^p}{s},$$

or

$$s \text{ and } s^{p-1} - ps^{p-2}x + \frac{p \cdot p-1}{\underline{|2}} s^{p-3}x^2 \ldots + \frac{p \cdot p-1}{\underline{|2}} sx^{p-2} + px^{p-1}.$$

Also s is prime to x, whence the result, which is due to Jaquemet (1651—1729), follows.

The second fact is that $\dfrac{x^p + y^p}{x + y}$ can be written in the form

$$\tfrac{1}{4}\left(U^2 - (-1)^{\frac{p-1}{2}} p V^2\right),$$

where U and V are polynomials in x and y, but the proof is more complicated than that for the first fact so it may be omitted here. In the particular case, however, when $n = 5$ or 7, there is no difficulty in finding U and V by elementary algebra.

Taking now the case $n = 5$, we have

$$z^5 = (x + y)\left(\frac{x^5 + y^5}{x + y}\right),$$

or

$$z^5 = (x + y)\left(\frac{U^2 - 5V^2}{4}\right),$$

where U and V are quadratic functions of x and y. From the above, the two factors on the right-hand side are either prime to each other, or have a common factor 5 of which the first power only will divide $\dfrac{U^2 - 5V^2}{4}$. We then have an equation of the form

$$\tfrac{1}{4}(U^2 - 5V^2) = W^5 \text{ or } 5W^5.$$

The difficulty arising in the case $n = 3$, and overlooked by Euler, occurs in the discussion of these equations, but it is possible to avoid it by similar methods. It follows also that the case $n = 5$ is impossible. A similar method applies to the case $n = 7$, but more algebra is required than for $n = 5$.

CHAPTER II

KUMMER'S WORK

The difficulties arising with increasing values of n soon made it clear that other methods were required for the general case. These, of which we shall now give an account, were introduced by Kummer (1810—1893). His results were the most important contribution to this subject by any mathematician either before or after his time. Not only were they the most general, in that he succeeded in proving Fermat's Last Theorem for a large number of values of n included in several classes, but they were also the most useful, and marked an important stage in the development of mathematics. The theory of ideals, which is now part of the fundamental groundwork of the Theory of Numbers, had its origin in Kummer's researches on this subject and the general law of reciprocity. His methods and results were the starting point of numerous investigations commenced many years after his time, and have led to some very surprising results even within the last twelve years. His work is an excellent illustration of the great indebtedness of mathematics and mathematicians to the consideration of one or two isolated questions.

Writing the equation (p an odd prime)

$$x^p + y^p = z^p$$

in the form

$$(x + y)(x + \zeta y)(x + \zeta^2 y) \dots (x + \zeta^{p-1} y) = z^p \quad \dots\dots\dots(6),$$

where ζ is a complex pth root of unity, the attention of mathematicians was drawn to the study of expressions of the form

$$a + b\zeta + c\zeta^2 + \dots k\zeta^{p-1},$$

where $a, b, c \dots$ are integers, and to inquire if the ordinary laws of arithmetic applied to such expressions.

Many of the most important developments of arithmetic depend upon the definition of a prime number and the so called factor theorem, namely that every number can be resolved into prime factors in one way only. It follows from this fact that if positive integers $A, B, C \dots K, L$, of which no two have a common factor, satisfy the condition

$$ABC \dots K = L^p,$$

then each of the numbers A, $B \dots K$ must be a perfect pth power. Should any of the quantities A, $B \dots$ have a common factor, this result must be slightly modified; for example A now will be a perfect pth power multiplied by a constant depending on the common factors mentioned above. Particular cases of this theorem have already been used. The question immediately suggests itself—Can this theorem be extended to apply to equation (6), and can we deduce that the factors $x+\zeta y$, $x+\zeta^2 y \dots$ are each pth powers of expressions of the form

$$a+b\zeta+c\zeta^2+\dots$$

or perhaps multiples of such pth powers? If so, a proof of Fermat's Last Theorem would be fairly easy.

ARITHMETICAL PROPERTIES OF NUMBERS OF THE FORM $a+ib$

Before we answer this question, let us consider what occurs in some analogous but simpler cases. The simplest case would be the study of complex numbers of the form

$$a+ib,$$

where $i=\sqrt{-1}$ and a and b are rational numbers. When a and b are integers it seems natural to call the complex number $a+ib$ a complex integer. When $b=0$ the complex integer becomes an ordinary integer. Further it is obvious that the sum, difference or product of two complex integers is also a complex integer, so that the definition of a complex integer is consistent.

As regards division, a complex integer $a+ib$ is said to be divisible by a complex integer $c+id$ if a complex integer $x+iy$ can be found so that

$$a+ib=(c+id)(x+iy).$$

We note that while 1 is exactly divisible by only two integers, namely ± 1, it is exactly divisible by four integers in the complex theory, namely ± 1, $\pm i$. The divisors of unity are called units.

The question now arises, "What is the definition of a prime number in the new theory?" The odd primes

$$3,\ 5,\ 7,\ 11,\ 13,\ 17,\ \dots$$

can be divided into two groups such as

$$5,\ 13,\ 17,\ 29,\ 37,\ \dots$$

of which every one leaves the remainder 1 when divided by 4; and

$$3,\ 7,\ 11,\ 19,\ \dots$$

every one of which leaves the remainder 3 when divided by 4.

The numbers in the first group, however, are no longer primes in the complex theory. For it is clear that

$$5 = (2 + i)(2 - i),$$
$$13 = (3 + 2i)(3 - 2i),$$
$$17 = (4 + i)(4 - i),$$

and it can be shown that every prime number of the form $4n + 1$ can be expressed in this way, that is

$$4n + 1 = (a + ib)(a - ib) = a^2 + b^2,$$

where a and b are integers. This fact was indeed stated by Fermat and first proved by Euler, but it is of an entirely different kind from the theorems of elementary arithmetic.

The numbers in the second group cannot be factorised in this way, for then

$$4n + 3 = (a + ib)(a - ib) = a^2 + b^2.$$

But as already remarked, the square of any integer when divided by 4 leaves a remainder 0 or 1. Hence $a^2 + b^2$ when divided by 4 can only leave a remainder 0, 1, 2 and not 3. This proves the statement.

The behaviour of the even prime 2 is very different from that of the odd primes. For

$$2 = i(1 - i)^2,$$

so that 2 is practically* a square number in the complex theory.

We can now define the prime numbers of the complex theory. These are the numbers

$$3,\ 7,\ 11,\ 19,\ \ldots,$$

that is the primes of the form $4n + 3$ in the ordinary theory; the complex quantities

$$2 \pm i, \quad 3 \pm 2i \ldots a \pm ib,$$

which are the factors of 5, 13, ... and of the primes of the form $4n + 1$; and lastly $1 - i$.

The factor theorem for primes in the new theory

It can be shown that in the complex theory, the primes, as just defined, have properties practically identical with those of the ordinary primes. For example every complex integer can be resolved into prime factors in one way and only one way, noting of course that the factors

$$a + ib, \quad -(a + ib), \quad \pm(a + ib)$$

are not considered as different.

* That is, except for the unit factor i.

Suppose for instance that $a + ib$ is a factor of the number p which is a prime in the ordinary theory, so that

$$a^2 + b^2 = p.$$

Then $x + iy$ is divisible by $a + ib$ if

$$\frac{x + iy}{a + ib} = \frac{(x + iy)(a - ib)}{a^2 + b^2}$$

is a complex integer. Hence

$$\xi = ax + by \quad \text{and} \quad \eta = ay - bx$$

must both be divisible by p. But as

$$a\xi - b\eta = (a^2 + b^2)\, x = px,$$

and a and b are both prime to p, it is clear that ξ and η are both divisible by p if one of them is. Hence the condition that*

$$(x_1 + iy_1)(x_2 + iy_2) \equiv 0 \pmod{(a + ib)},$$

is that $\quad b(x_1x_2 - y_1y_2) - a(x_1y_2 + x_2y_1) \equiv 0 \pmod{p},$

or multiplying by b

$$b^2(x_1x_2 - y_1y_2) - ab(x_1y_2 + x_2y_1) \equiv 0 \pmod{p},$$

which since $\quad a^2 + b^2 \equiv 0 \pmod{p},$

can be written as

$$(ay_1 - bx_1)(ay_2 - bx_2) \equiv 0 \pmod{p}.$$

Hence one of these two factors must be divisible by p, that is, the corresponding factor $x_1 + iy_1$ or $x_2 + iy_2$ is divisible by p. This shows that practically the same arithmetical laws hold for complex integers as for ordinary ones.

THE EQUATION $x^2 + y^2 = z^n$

We can now solve the equation

$$x^2 + y^2 = z^n \quad \text{.........................}(6\,a),$$

where x, y and z are ordinary integers no two of which have a common factor, by writing

$$(x + iy)(x - iy) = z^n.$$

Now in the complex theory x and y are still prime to each other, so that the common factor of the two complex quantities $x + iy$, $x - iy$ must be a divisor of 2, that is the common factor must be 1, $1 + i$ or 2. We can exclude 2 because then x and y would both be even.

We can also exclude $1 + i$, for if

$$\frac{x + iy}{1 + i} = \frac{x + y + i(y - x)}{2}$$

* The statement $A \equiv B \pmod{C}$ means that $A - B$ is divisible by C.

were a complex integer, $x+y$ and $x-y$ would both be divisible by 2, or since x and y are both prime to each other, this means that x and y are both odd. But this is impossible for then x^2+y^2 would be double an odd number and could not be a perfect nth power if $n>1$. Hence as $x+iy$ and $x-iy$ are prime to each other, each of them, except for a unit factor, must be a perfect nth power, so that the solution of the equation (6 a) is given by

$$x+iy=i^r(a+ib)^n, \qquad x-iy=i^{-r}(a-ib)^n,$$
$$z=a^2+b^2,$$

where a and b are ordinary integers such that x, y and z have no common factor, and r is any integer.

Numbers of other kinds, and their factorisation

The difficulties arising then in the discussion of complex integers of the type $a+ib$ are comparatively simple. These complex integers naturally suggest algebraic integers of a more general type, such as for example, $a+b\sqrt{m}$, where a, b and m are rational quantities. We shall consider in particular the study of the quantities $x=a+b\sqrt{-5}$, where a and b are rational, and shall call x an algebraic integer if a and b are integers. A more general definition is that x is an algebraic integer if it is a root of an algebraic equation of which the coefficients are integers, while the coefficient of the highest power of x is unity. The two definitions are equivalent in our special case, though they would not be so if $\sqrt{-5}$ were replaced by $\sqrt{5}$. The addition, subtraction, division or multiplication of such integers calls for no comment. It is when we start to factorise such algebraic integers that a difficulty soon arises. It seems natural to call an algebraic integer $x+y\sqrt{-5}$ a prime when it is not divisible by $a+b\sqrt{-5}$ unless $a=\pm x$, $b=\pm y$. Let us accept this definition of a prime.

Take the number 21 for example. Clearly we have

$$21=3\times 7=(4+\sqrt{-5})(4-\sqrt{-5})=(1+2\sqrt{-5})(1-2\sqrt{-5}).$$

But $4+\sqrt{-5}$ cannot be resolved into a product of factors of algebraic integers of the form $a+b\sqrt{-5}$. For if this were possible, then

$$4+\sqrt{-5}=(a+b\sqrt{-5})(c+d\sqrt{-5}),$$

say, so that $\qquad 4-\sqrt{-5}=(a-b\sqrt{-5})(c-d\sqrt{-5}).$

Hence by multiplication

$$21=(a^2+5b^2)(c^2+5d^2),$$

so that a^2+5b^2, being a factor of 21, must be either 1, 3, 7 or 21. This gives $a=\pm 1$, $b=0$; or $a=\pm 4$, $b=\pm 1$; or $a=\pm 1$, $b=\pm 2$. The solution $a=\pm 1$, $b=0$ does not give a factor of $4+\sqrt{-5}$, while the solution $a=\pm 4$, $b=\pm 1$ does not give a factor, since

$$\frac{4+\sqrt{-5}}{4-\sqrt{-5}}=\frac{11+8\sqrt{-5}}{21}$$

is not a complex integer. It is also easily seen that $\pm 1 \pm 2\sqrt{-5}$ is not a factor.

In the same way it is found that neither $1+2\sqrt{-5}$, 3 nor 7 splits into factors of the form $a+b\sqrt{-5}$. We have then 21 expressed as a product of primes in three different ways. Moreover

$$(4+\sqrt{-5})(4-\sqrt{-5}) \text{ is divisible by } 3,$$

where 3 is a prime in the new theory, while $4+\sqrt{-5}$ and $4-\sqrt{-5}$ are both prime to 3. Hence the factor theorem of arithmetic which states that the product of two integers ab cannot be divisible by a prime p, unless either a or b is divisible by p, no longer applies. This breaking down of one of the fundamental laws of arithmetic for integers of the type $a+b\sqrt{-5}$ brings us face to face with a great difficulty, and suggests that the method of defining a prime in the present instance, in the same manner as for integers of the form $a+ib$, is not satisfactory. It is however not obvious at first sight how to suggest an alternative method. Unless this is done, we cannot deduce from the equation $xy=z^2$, where x and y are prime to each other, that x and y are both perfect squares.

For example $\qquad (2+\sqrt{-5})(2-\sqrt{-5})=9,$

where the factors $2\pm\sqrt{-5}$ are primes, have no common factor, and are not equal to the squares of integers of the form $a+b\sqrt{-5}$.

THE DIFFICULTY ARISING IN THE GENERAL EQUATION

It is now clear that given an equation of the form

$$(x+y)(x+\zeta y)(x+\zeta^2 y)\dots(x+\zeta^{p-1}y)=z^p \quad \dots\dots\dots(6),$$

where ζ is a complex pth root of unity, it cannot be asserted that $x+\zeta y$, for example, is the pth power of an expression of the form.

$$a+b\zeta+c\zeta^2+\dots k\zeta^{p-1},$$

until an investigation has been made of the arithmetical properties of such algebraic integers. It may happen that the definition of a prime in the new theory in the same way as for numbers of the form $a+ib$

will be satisfactory, in which case the algebraic integers can be factorised in one way only; or the same difficulty may arise as in the case of numbers of the form $a+b\sqrt{-5}$. As a matter of fact, it is not true that the algebraic numbers above can be factorised uniquely, but the first case of failure occurs when $p=23$. It is not surprising then that such mathematicians as Lamé, Cauchy and even Kummer should have been originally under the impression that the algebraic integers above could be factorised uniquely.

Lamé made this false assumption in giving a proof of Fermat's Last Theorem, as was pointed out by Liouville and Kummer. Kummer also had previously made the same mistake in attempting a proof, as was pointed out by Dirichlet, who expressed his belief that the algebraic numbers involved could not in general be factorised uniquely.

The question before us, then, is the removal of the difficulty mentioned above, and a very simple illustration due to Hilbert may show us how this can be done. Let us consider only the odd integers of the form $4n+1$, that is to say 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61, 65, 69 ..., and investigate what happens when we attempt to build up the arithmetical laws for this group of integers, say a, b, c .... A number a will now be called a prime if it cannot be expressed in the form $a=bc$ unless b or c equals unity. Thus 21 would be a prime number in the new theory, for although

$$21=3\times 7,$$

neither 3 nor 7 is included in the group of integers of the form $4n+1$. Also $49\times 9=21^2$, and neither 9 nor 49 is the square of a number of the group, while 9 and 49 are prime to each other, since no number* of the group divides both of them. Again 693 can be factorised as

$$693=9\times 77=21\times 33,$$

that is in two essentially different ways, since 9, 21, 33 and 77 must be considered as primes in the new theory.

The difficulties arising now are of exactly the same kind as arose in the consideration of algebraic numbers of the form $a+b\sqrt{-5}$. But the way out of the difficulty is obvious in the present case. Instead of considering only the integers 1, 5, 9, 13 ..., we consider in addition the odd numbers of the form $4n+3$, for example 3, 7, 11 .... Then we know that in the new group of integers, 1, 3, 5, 7, 9 ..., the ordinary laws of arithmetic hold, and now

$$9=3^2,\quad 77=7\times 11,\quad 21=3\times 7,\quad 33=3\times 11,$$

* Except unity.

so that the two different methods of factorising 693 reduce to the one way

$$693 = 3^2 \times 7 \times 11.$$

This method of removing the difficulty is very simple and general, and suggests at once the question—Can this idea be extended to the algebraic numbers of the form $a + b\sqrt{-5}$; or in other words can the group of algebraic numbers of the form $a + b\sqrt{-5}$ be enlarged by joining a new group of numbers, so that the factor law of arithmetic holds for this enlarged group? The answer is in the affirmative, not only for these special algebraic numbers, but for all algebraic numbers. The method of doing this has been presented in three different ways* by Kummer†, Dedekind, Kronecker and Weber. The principles underlying them are essentially the same, and are now included under what is known as the arithmetical theory of algebraic numbers.

Introduction of ideal numbers

The methods may be made clearer if presented in a rather different way from those of the investigators above, but which, though very useful for giving a clear insight into the matter, would prove rather difficult if made the starting point of an investigation for the general algebraic number.

For our purpose it may be sufficient to say that instead of considering algebraic numbers of the form $a + b\sqrt{-5}$ we consider the new group of numbers defined by τ, where

$$\tau^2 = x + y\sqrt{-5},$$

and x and y are any integers, whose greatest common factor is a perfect square or 5 times a perfect square, and satisfying the condition that

$$x^2 + 5y^2$$

should be a perfect square. The group of algebraic integers now arising reproduces itself by multiplication and its members may be called ideal numbers. It includes as part of itself the numbers of the form $a + b\sqrt{-5}$, of which we have already spoken.

The ordinary primes

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47 ...

can now be divided into three classes, while 5 is in a class by itself.

* See Bachmann, *Zahlentheorie*, Vol. v. p. 521, for two other methods by Hensel and Sochotzki.

† Kummer dealt only with the algebraic numbers arising from the complex roots of unity.

The first class consists of primes such as 29, 41, 61 ..., which can be written in the form

$$29 = (3 + 2\sqrt{-5})(3 - 2\sqrt{-5}),$$
$$41 = (6 + \sqrt{-5})(6 - \sqrt{-5}),$$
$$61 = (4 + 3\sqrt{-5})(4 - 3\sqrt{-5}),$$

that is, each of them can be expressed in the form $a^2 + 5b^2$.

The second class consists of primes such as 3, 7, 23 ..., which cannot be expressed in the form $a^2 + 5b^2$, though their squares can be expressed in this form, for example

$$3^2 = (2 + \sqrt{-5})(2 - \sqrt{-5}),$$
$$7^2 = (2 + 3\sqrt{-5})(2 - 3\sqrt{-5}),$$

and in general for any prime p of this kind

$$p^2 = c^2 + 5d^2,$$

or if ideal numbers τ_1, τ_2 are defined by

$$\tau_1^2 = c + d\sqrt{-5}, \quad \tau_2^2 = c - d\sqrt{-5},$$

then p can be factorised in the form

$$p = \tau_1 \tau_2.$$

The third class of primes consists of primes, say q, such as

$$2,\ 11,\ 13,\ 17\ \ldots,$$

which are such that neither they nor their squares can be expressed* in the form $a^2 + 5b^2$. A very simple rule enables us to distinguish between the three classes of primes, but the principles employed depend upon the Theory of Numbers. The prime 5 as already remarked is in a class by itself and since $5 = -(\sqrt{-5})^2$, 5 is practically a square number.

The proof of a unique factorisation law

It can be shown that the new group of algebraic integers of the type specified by $\tau = \sqrt{x + y\sqrt{-5}}$, and which includes the complex numbers of the form $m + n\sqrt{-5}$ as a special case, can be factorised uniquely by means of the quantities just denoted by $a + b\sqrt{-5}$, $\sqrt{c + d\sqrt{-5}}$, q and $\sqrt{-5}$.

This follows from the condition of divisibility of an ideal number by the new primes, just as in the case of complex numbers of the form $x + iy$. For example, if the ideal number $\sqrt{x + y\sqrt{-5}}$ is divisible by $\sqrt{2 + \sqrt{-5}}$ say τ, so that τ is an ideal factor of 3, we must have

$$\sqrt{x + y\sqrt{-5}} = \sqrt{2 + \sqrt{-5}}\sqrt{m + n\sqrt{-5}},$$

* We assume that $b \neq 0$.

where m and n are integers. It is easily seen that

$$9m = 2x + 5y, \quad 9n = -x + 2y,$$

and that both m and n are integers if

$$x - 2y \equiv 0 \pmod 9.$$

Similarly the number $x + y\sqrt{-5}$ is divisible by τ if

$$\sqrt{x^2 - 5y^2 + 2xy\sqrt{-5}} \text{ is divisible by } \tau.$$

The condition for this is that

$$x^2 - 5y^2 - 4xy \equiv 0 \pmod 9,$$

that is $$(x - 2y)^2 \equiv 0 \pmod 9,$$

or $$x - 2y \equiv 0 \pmod 3.$$

We can now prove that if the product of two algebraic numbers is divisible by τ, one of them is divisible by τ. Let the numbers be

$$x_1 + y_1\sqrt{-5}, \quad x_2 + y_2\sqrt{-5}.$$

Their product

$$x_1x_2 - 5y_1y_2 + \sqrt{-5}\,(x_1y_2 + x_2y_1)$$

being divisible by τ, we must have

$$x_1x_2 - 5y_1y_2 - 2(x_1y_2 + x_2y_1) \equiv 0 \pmod 3,$$

that is $$(x_1 - 2y_1)(x_2 - 2y_2) \equiv 0 \pmod 3.$$

This means that one of these two factors must be divisible by 3, so that, if say,

$$x_1 - 2y_1 \equiv 0 \pmod 3,$$

then from the above $x_1 + y_1\sqrt{-5}$ must be divisible by τ.

The same result follows if we consider the product of two ideal numbers $\sqrt{x_1 + y_1\sqrt{-5}}$, $\sqrt{x_2 + y_2\sqrt{-5}}$. This product is divisible by τ if

$$x_1x_2 - 5y_1y_2 - 2(x_1y_2 + x_2y_1) \equiv 0 \pmod 9,$$

or $$(x_1 - 2y_1)(x_2 - 2y_2) \equiv 0 \pmod 9,$$

and a simple discussion shows that if $x_1 - 2y_1$ is divisible by 3, it is also divisible by 9 (noting that $x_1^2 + 5y_1^2$ is a square). Hence as before one of the ideal numbers must be divisible by τ.

It is clear now that being given an algebraic number $x + y\sqrt{-5}$, the condition

$$x - 2y \equiv 0 \pmod 3$$

is sufficient to define the ideal prime factor $\sqrt{2 + \sqrt{-5}}$ of the complex number, and that the actual form of the ideal number need not be

given explicitly. Dedekind, for example, put $x-2y=3m$, so that $x+y\sqrt{-5}$ becomes $3m+y(2+\sqrt{-5})$, and then considered the properties of the groups of numbers arising by taking different values for m and y, and called the group of numbers an ideal. Kronecker however would have studied the linear expression as a function of x, y; while Kummer would have used the congruence

$$x-2y\equiv 0 \pmod 3$$

as defining the ideal prime τ.

Application of ideal numbers to Fermat's Last Theorem

For algebraic numbers of the form

$$a+b\zeta+c\zeta^2+\dots$$

Kummer showed that the ideal numbers were of the form

$$\sqrt[r]{a_1+b_1\zeta+c_1\zeta^2\dots},$$

where a_1, b_1, c_1, ... are integers satisfying certain conditions, while r is a factor of a number called the number of ideal classes. Its value depends only on p, and can be found by a very complicated method depending on principles introduced into analysis by Dirichlet.

Continuing now the discussion of the equation

$$(x+y)(x+\zeta y)\dots(x+\zeta^{p-1}y)=z^p \quad \dots\dots\dots\dots\dots(6),$$

consider first the case when z is prime to p. This is tantamount to saying that no two of the factors on the left-hand side have a common factor. Hence by introducing ideal numbers, we have, practically as in the case of the equation $x^2+y^2=z^m$, a number of equations of the form

$$x+\zeta y=\xi\tau^p,$$
$$x+\zeta^2 y=\xi_2\tau_2^{\,p},$$
$$\dots\dots\dots\dots\dots$$

where τ, τ_2, ... are ideal numbers and ξ, ξ_2, ... are units, i.e. quantities of the form

$$a_2+b_2\zeta+c_2\zeta^2\dots,$$

which are divisors of unity.

Noting now the explicit expression for τ, namely

$$\tau=\sqrt[r]{a+b\zeta+c\zeta^2\dots},$$

we have
$$x+\zeta y=\xi\left(\sqrt[r]{a+b\zeta+c\zeta^2\dots}\right)^p.$$

If now r is prime to p, and this is an extremely important condition, it follows that

$$a+b\zeta+c\zeta^2\dots$$

is the rth power of a similar expression, so that we can write

$$x + \zeta y = \xi\,(\alpha + \beta\zeta + \gamma\zeta^2 \ldots)^p,$$

where α, β, γ, ... are integers. We find other equations by changing ζ into ζ^2, ζ^3, ... It is then a comparatively simple matter to show that equations of this kind are impossible, not only when x and y are ordinary integers but also when they are integers of the form

$$A + B\zeta + C\zeta^2 + \ldots.$$

A similar conclusion can be drawn in the case when z is not prime to p. Hence Fermat's Last Theorem is proved in all the cases where r or the number of classes of ideals is prime to p. The condition for this can be stated in a remarkable form by noting the following expansion in ascending powers of x, namely

$$\frac{x}{e^x - 1} = 1 - \frac{x}{2} + \sum_{n=1}^{\infty} (-1)^{n-1} \frac{B_n x^{2n}}{(2n)!},$$

so that $B_1 = \frac{1}{6}$, $B_2 = \frac{1}{30}$, $B_3 = \frac{1}{42}$, ... are the well known Bernoulli's numbers. Then the required condition is that the numerators of none of the first $\frac{1}{2}(p-3)$ of the Bernoulli's numbers should be divisible by p. The only primes less than 100 for which this condition is not satisfied are $p = 37$, 59, 67, and hence it is proved* that

$$x^p + y^p = z^p$$

is impossible if p is an odd prime less than 100, except when $p = 37$, 59, 67.

In order to establish the truth of the theorem for these exceptional values of p, Kummer gave in 1857 some additional results for primes satisfying certain conditions†. These conditions were satisfied by $p = 37$, 59, 67, so that Fermat's Last Theorem is proved for *all* values of p, prime or otherwise, less than 100, omitting of course $p = 2$.

ANOTHER RESULT BY KUMMER

Some other important consequences were deduced by Kummer in the special case when one of the unknowns is not divisible by p. We

* A complete account of Kummer's theory of ideal numbers is given in *Liouville's Journal*, t. XVI. 1851. Hilbert in his well known report on "Die Theorie der algebraischen Zahlkörper" gives the modern version. The French translation of this report has an appendix giving other results on Fermat's Last Theorem. For a good introduction, see Sommer, *Vorlesungen über Zahlentheorie*.

† It appears that Kummer has made some errors which vitiate his proof for the cases $p = 37$, 59, 67. See Vandiver "On Kummer's memoir of 1857 concerning Fermat's Last Theorem," *Proceedings of the National Academy of Science, Washington, U.S.A.*, Vol. VI. May, 1920. The case $p = 37$, however, was proved impossible by Mirimanoff in 1892, so that the cases $p = 59$, 67 are still doubtful.

saw previously that we could deduce from equation (6) a series of equations of the form

$$x + \zeta^r y = \xi_r \tau_r^p,$$

where $r = 1, 2, \dots p$. Kummer showed that it was possible to select $\frac{p-1}{2}$ of these equations in such a way that the product of the corresponding τ_r's is an actual number of the form

$$T = a + b\zeta + c\zeta^2 + \dots.$$

By multiplying together the product of the $\frac{p-1}{2}$ equations, he obtained a result of the form

$$\Pi\,(x + \zeta^r y) = \pm\,\zeta^k T^p,$$

where the multiplication on the left refers to the selected $\frac{p-1}{2}$ factors. Replacing ζ by e^v, we have an identity of the form

$$\Pi\,(x + e^{rv} y) = \pm\,e^{kv}(a + be^v + \dots)^p + (1 + e^v + e^{2v} + \dots e^{(p-1)v}) f(v),$$

where $f(v)$ is a polynomial in e^v with integral coefficients. By differentiation he deduced* that

$$B_n \frac{d^{p-2n}}{dv^{p-2n}}[\operatorname{Log}(x + e^v y)]_{v=0} \equiv 0 \pmod p,$$

when $n = 1, 2, \dots \tfrac{1}{2}(p-3)$, where B_n is the nth Bernoullian number as defined before, a result which has since proved very useful.

Kummer, although not a candidate for a prize offered by the French Academy for a proof of Fermat's Last Theorem, was awarded it in recognition of his researches on complex numbers. Certainly never was an investigator on these subjects more worthy of one.

Deductions from Kummer's last result

For about fifty years after Kummer's work, very little was accomplished either in extending or in developing the full consequences of his results on Fermat's Last Theorem. In the early part of this century, however, mathematicians turned once more to Kummer's results, and in particular to the one, that if

$$x^p + y^p = z^p$$

had solutions for which z is prime to p, then

$$B_n \frac{d^{p-2n}}{dv^{p-2n}}[\operatorname{Log}(x + e^v y)]_{v=0} \equiv 0 \pmod p$$

for $n = 1, 2, 3, \dots \tfrac{1}{2}(p-3)$.

* *Abhand. Ak. Wiss. Berlin*, 1857.

This result, by putting

$$p - 2n = i,$$

can be written in the slightly different form

$$B_{\frac{1}{2}(p-i)} \frac{d^i}{dv^i} [\text{Log}(x + e^v y)]_{v=0} \equiv 0 \pmod{p} \quad \ldots\ldots\ldots(7),$$

where $i = 3, 5, \ldots p - 4, p - 2.$

Take now $i = 3$, then

$$B_{\frac{1}{2}(p-3)} \frac{d^3}{dv^3} [\text{Log}(x + e^v y)]_{v=0} \equiv 0 \pmod{p},$$

and this reduces to

$$B_{\frac{1}{2}(p-3)}\, xy(x - y) \equiv 0 \pmod{p}.$$

Hence if $B_{\frac{1}{2}(p-3)}$ is not divisible by p,

$$xy(x - y) \equiv 0 \pmod{p}.$$

If then the equation has a solution for which x, y, z are all prime to p,

$$x \equiv y \pmod{p};$$

and in exactly the same way

$$x \equiv z \pmod{p}.$$

This gives $3x^p \equiv 0 \pmod{p}$,

which is impossible if p is not equal to 3.

Hence we have proved the result that the equation

$$x^p + y^p = z^p$$

cannot have integer solutions for which x, y and z are all prime to p unless $B_{\frac{1}{2}(p-3)}$ is divisible by p. In the same way, by taking $i = 5, 7, 9$, it is found that in addition

$$B_{\frac{1}{2}(p-5)}, \quad B_{\frac{1}{2}(p-7)}, \quad B_{\frac{1}{2}(p-9)}$$

must be divisible by p. This was proved by Mirimanoff in 1905, for the last two of the four Bernoulli's numbers above, and by Kummer for the first two, but the case for $B_{\frac{1}{2}(p-3)}$ had been practically announced previously by Cauchy, although without proof.

Mirimanoff also showed by developing the value of

$$\frac{d^i}{dv^i} [\text{Log}(x + e^v y)]_{v=0},$$

that, if x, y, z are all prime to p, then Kummer's result (7) could be expressed in the form*

$$B_{\frac{1}{2}(p-i)} (t - 2^{i-1} t^2 + 3^{i-1} t^3 \ldots \pm (p-1)^{i-1} t^{p-1}) \equiv 0 \pmod{p},$$

* *Crelle's Journal*, Vol. CXXVIII.

or say
$$B_{\frac{1}{2}(p-i)}\,\phi_i(t) \equiv 0 \pmod{p},$$
where
$$i = 3,\ 5,\ \dots\ p-2,$$
and t is the ratio of any pair of x, y, z. Or again in the form
$$\phi_n(t)\,\phi_{p-n}(t) \equiv 0 \pmod{p}$$
where
$$n = 1,\ 2,\ \dots\ p-1.$$

The congruences $2^{p-1} \equiv 1 \pmod{p^2}$, $3^{p-1} \equiv 1 \pmod{p^2}$ ....

A number of conditions can be found by eliminating t from the above congruences. Although there seems no *a priori* reason for expecting simple results, Wieferich* showed in 1909 that one of these conditions could be expressed in the surprisingly simple form
$$2^{p-1} \equiv 1 \pmod{p^2}.$$

This extremely simple and unexpected result represented the first real advance made in the subject since Kummer's work. The Göttingen Academy of Science awarded him 100 marks from the interest of the Wolfskell fund.

In 1913, Meissner showed that $p = 1093$ was the only prime less than 2000 for which this congruence was satisfied.

In other words the equation (p a prime)
$$x^p + y^p = z^p,$$
where
$$2000 > p > 2,$$
cannot be satisfied by values of x, y and z, each of which is prime to p, except perhaps when $p = 1093$.

Simpler proofs of Wieferich's result were soon given by Frobenius and Mirimanoff, the latter also showing† that under the same conditions
$$3^{p-1} \equiv 1 \pmod{p^2}.$$

The two congruences above could not of course have been foreseen from Kummer's original results, but another proof was given by Furtwängler, which seems more natural and simple, depending upon ideas which should be capable of further extension. The following may perhaps give some indication of the ideas involved.

Suppose we have two odd primes, say 3 and 7, and it is required to investigate if integers x and y can be found so that
$$x^2 \equiv 7 \pmod{3}, \quad y^2 \equiv 3 \pmod{7}.$$

* *Crelle's Journal*, Vol. CXXXVI.

† *Crelle's Journal*, Vol. CXXXIX.

It is easy to see that the first congruence is satisfied by $x = 2$, while the second congruence is impossible. If however 3 and 7 had been replaced by any odd two primes p and q, a very simple theorem known as the law of quadratic reciprocity enables us from the known possibility or impossibility of one of the congruences to determine if the other congruence is possible or not. But if one of the congruences is impossible, we can at once conclude that the equation

$$x^2 = py^2 + qz^2$$

is also impossible.

This theorem, moreover, can be extended to congruences of the form

$$x^p \equiv P \pmod{Q},$$

where x, P, Q are algebraic numbers of the form

$$a + b\zeta + c\zeta^2 + \dots,$$

or even ideal numbers occurring in the theory of such algebraic numbers. The theorem in this case, known as the general law of reciprocity, was enunciated and proved by Kummer, but only a special case, due to Eisenstein, was required in Furtwängler's* proof. Assuming this result, Furtwängler's proof of the results

$$2^{p-1} \equiv 3^{p-1} \equiv 1 \pmod{p^2}$$

is very simple and natural. As already remarked, it seems that a new application of the laws of reciprocity may be expected to lead to interesting results.

It appears probable that the 2 and 3 above can be replaced by any prime q (except p). In 1914, Vandiver† showed that 5 was another value for q, while Frobenius‡ showed that q might also take the values 11 and 17; and also the values 7, 13, 19, if $p \equiv 5 \pmod{6}$.

The proofs, however, are very complicated and depend upon a special study of the properties of Bernoulli's numbers. The elimination is carried out by taking the congruence

$$B_{\frac{1}{2}(p-i)} \phi_i(t) \equiv 0 \pmod{p},$$

multiplying throughout by an appropriate function of t, say $f_i(t)$, and then adding together the left-hand sides of the congruences, which then reduce practically to the form

$$q^{p-1} \equiv 1 \pmod{p^2},$$

for the values $q = 2, 3, 5 \dots$ as just noted.

* *Sitzungs. Ak. Wiss. Wien* (Math.), Vol. CXXI. 1912 II *a*, pp. 589—592.

† *Crelle's Journal*, Vol. CXLIV. 1914, p. 314.

‡ *Sitzungs. Ak. Wiss. Berlin*, 1914, p. 653.

CHAPTER III

LIBRI'S RESULT

We shall now pass on to other methods* which have been employed. These, although of interest, do not prove the truth of Fermat's Last Theorem for even one case.

A simple method of attempting to prove the Theorem, which soon suggests itself to investigators, may be explained by taking the particular case

$$x^3 + y^3 + z^3 = 0,$$

which has already been considered. It follows from this equation that one of the unknowns must be divisible by 3; for otherwise each of them would be of the form $3n \pm 1$, and then their cubes would be of the form $27n^3 \pm 27n^2 + 9n \pm 1$, that is of the form $9m \pm 1$. But obviously the sum of three numbers each of the form $9m \pm 1$ cannot be zero, as this sum is not divisible by 9. Hence one of the unknowns must be divisible by 3.

Similarly it can be shown that one of the unknowns must be divisible by 7. For it is easily shown that the cubes of all numbers not divisible by 7 are of the form $7m \pm 1$, so that the sum of the cubes of three numbers cannot be divisible by 7, and hence certainly not equal to zero, unless one of the numbers is divisible by 7.

The question at once arises—Can an infinite number of primes q be found with the same property as the primes 3 and 7 above; that is to say, from the fact that $x^3 + y^3 + z^3$ is divisible by q, does it follow that one of the unknowns must be divisible by q? If so, the equation will be impossible, since one of the unknowns will be divisible by an infinite number of primes. A similar question suggests itself for the equation

$$x^p + y^p + z^p = 0.$$

Libri in 1832 stated without proof that an infinite number of primes such as q did not exist. This was proved by Pellet about 1886, and independently in 1909 by Dickson and Hurwitz amongst others. Dickson † also showed that

$$x^p + y^p + z^p$$

* Bachmann, *Niedere Zahlentheorie*, Vol. II. Chapter IX. will be found useful in connection with the first and third chapters of this book.

† *Crelle's Journal*, Vol. cxxxv. 1909, p. 181. Cf. also the paper by Hurwitz in Vol. cxxxvi. p. 272. A simple and elementary proof with rather larger limits for q

could be made divisible by q without any of the quantities x, y, z being divisible by q if

$$q \geqslant (p-1)^2(p-2)^2 + 6p - 2.$$

Hence the suggested method of attack cannot succeed in proving the truth of Fermat's Last Theorem.

Sophie Germain's result

Another line of attack depends upon some formulae discovered independently by a number of investigators, among whom may be mentioned Legendre, Abel and Peter Barlow, and developed by others such as Sophie Germain. It may be noted that Barlow was the first Englishman to write a treatise on the Theory of Numbers, and was also amongst the earliest writers who have given erroneous proofs of Fermat's Last Theorem.

Instead of $$x^p + y^p = z^p,$$
consider the more symmetrical form

$$x^p + y^p + z^p = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots(8),$$

which can be written as

$$-z^p = (x+y)\left(\frac{x^p+y^p}{x+y}\right).$$

We note now* that either z is not divisible by p, in which case the two factors on the right-hand side are prime to each other, or that z is divisible by p, in which case the two factors have a common factor p of which the first power only is contained in

$$\frac{x^p+y^p}{x+y}.$$

It follows now from symmetry, that if x, y and z are all prime to p, then

$$y+z = a^p, \quad \frac{y^p+z^p}{y+z} = \xi^p, \quad x = -a\xi,$$

$$z+x = b^p, \quad \frac{z^p+x^p}{z+x} = \eta^p, \quad y = -b\eta,$$

$$x+y = c^p, \quad \frac{x^p+y^p}{x+y} = \zeta^p, \quad z = -c\zeta,$$

from which

$$2x = b^p + c^p - a^p, \quad 2y = c^p + a^p - b^p, \quad 2z = a^p + b^p - c^p.$$

(prime or composite) was given by Schur in the *Jahresber. d. Deutschen Math.-Vereinigung*, Vol. xxv. 1916.

* See p. 8.

If however one of the unknowns, say z, is divisible by p, the third of the above group of equations must be replaced by

$$x + y = p^{p-1} c^p, \quad \frac{x^p + y^p}{x + y} = p\zeta^p, \quad z = -cp\zeta.$$

It is in the first case only, however, that important practical consequences have been deduced. Suppose it is possible to find an odd prime q satisfying the two following conditions: firstly, that the congruence

$$x^p + y^p + z^p \equiv 0 \pmod{q}$$

requires that one of x, y, z must be divisible by q, and secondly that no integer k can be found satisfying

$$k^p - p \equiv 0 \pmod{q};$$

then the equation (8) cannot be satisfied by integers x, y and z each of which is prime to p.

For since $$x^p + y^p + z^p = 0,$$
and is hence divisible by q, it follows that one of x, y, z must be divisible by q, say x. Now since

$$2x = b^p + c^p + (-a)^p$$

is divisible by q, one of a, b, c must be divisible by q. But b cannot be divisible by q, for then

$$y = -b\eta$$

would be divisible by q, contrary to the hypothesis that x and y are prime to each other. Similarly c cannot be divisible by q, so that a is divisible by q. Hence

$$y + z = a^p$$

is divisible by q. As x is divisible by q and

$$\eta^p = \frac{z^p + x^p}{z + x},$$

it follows that $$\eta^p \equiv z^{p-1} \pmod{q}.$$

Moreover from

$$\xi^p = \frac{y^p + z^p}{y + z} = y^{p-1} - y^{p-2} z \ldots + z^{p-1},$$

and $$y + z \equiv 0 \pmod{q}, \text{ i.e. } z \equiv -y \pmod{q},$$

it follows that

$$\xi^p \equiv pz^{p-1} \pmod{q},$$

or $$\xi^p \equiv p\eta^p \pmod{q}.$$

But η is prime to q since $y = -b\eta$, so that we can find an integer k to satisfy

$$\xi \equiv k\eta \pmod{q},$$

from which $$k^p \equiv p \pmod{q}.$$

But by hypothesis such an integer k does not exist. This proves then that equation (8) cannot be satisfied by integers each of which is prime to p. For example if $p = 7$, i.e.

$$x^7 + y^7 + z^7 = 0,$$

we can take $q = 29$. For the seventh powers of all numbers prime to 29 leave remainders ± 1, ± 12 when divided by 29. Hence from

$$x^7 + y^7 + z^7 \equiv 0 \pmod{29},$$

it follows that one of x, y, z must be divisible by 29. Also no number k can be found such that

$$k^7 \equiv 7 \pmod{29}.$$

Hence the equation can only have solutions for which one of x, y, z is divisible by 7.

The general theorem above is due to Sophie Germain, who gave the corresponding prime q for all odd primes p less than 100.

Wendt's form of the result

Sophie Germain's general theorem was given in a slightly different form by Wendt* in 1895. He also gave the following necessary and sufficient conditions for the existence of a prime typified by q in the preceding chapter. Firstly q must be of the form

$$q = 2hp + 1,$$

where h is an integer prime to 3, for otherwise the congruence

$$x^p + y^p + z^p \equiv 0 \pmod{q}$$

would have solutions† for which x, y, z are all prime to q.

Secondly q must not be a divisor of the determinant

$$D_{2h} = \begin{vmatrix} \binom{2h}{1}, & \binom{2h}{2}, & \cdots & \binom{2h}{2h-1}, & 1 \\ \binom{2h}{2}, & \binom{2h}{3}, & \cdots & 1, & \binom{2h}{1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1, & \binom{2h}{1}, & \cdots & \binom{2h}{2h-2}, & \binom{2h}{2h-1} \end{vmatrix}.$$

* *Crelle's Journal*, Vol. CXIII. 1894, p. 335.

† If $h = 3n$, we could take $x = 1$, $y = g^{2n}$, $z = g^{4n}$, where g is a primitive root (mod q).

This second condition was found by considering the condition that the congruence

$$x^p + y^p + z^p \equiv 0 \pmod{q} \quad \text{.....................(9)}$$

should have solutions in which no one of x, y, z is divisible by q.

Putting in (9) $\quad x \equiv uz, \quad y \equiv -vz,$

we have $\quad u^p + 1 \equiv v^p \pmod{q} \quad$ ....(10),

where neither u nor v is divisible by q. Hence by a very well known theorem also due to Fermat, since

$$q = 2hp + 1,$$

$$\left.\begin{array}{l} u^{2hp} \equiv 1 \pmod{q} \\ v^{2hp} \equiv 1 \pmod{q} \end{array}\right\} \quad \text{.........................(10).}$$

By eliminating u, v between the last three congruences, it can be shown that the necessary and sufficient condition that the congruence (9) can be satisfied by values of x, y, z each of which is prime to q is that q should be a divisor of the determinant D_{2h}.

Wendt also replaced Sophie Germain's condition

$$k^p \not\equiv p \pmod{q}$$

by $\quad p^{2h} \not\equiv 1 \pmod{q}.$

From his determinant condition it would be extremely difficult to prove that a prime q can be found for a given value of p, especially as from Libri's theorem it is known that there cannot be more than a finite number of values of q for a given prime p. It can however be shown that if any one of the numbers

$$2p + 1, \quad 4p + 1, \quad 8p + 1, \quad 16p + 1, \quad 10p + 1, \quad 14p + 1$$

is a prime, it can be taken as a value for q. Suppose for example that

$$q = 2p + 1$$

is a prime. Then

$$D_2 = \begin{vmatrix} 2, & 1 \\ 1, & 2 \end{vmatrix} = 3,$$

so that q is prime to D_2. The other condition

$$p^2 \not\equiv 1 \pmod{q},$$

that is $\quad (p + 1)(p - 1) \not\equiv 0 \pmod{(2p + 1)},$

is also satisfied. Hence if $2p + 1$ is a prime, e.g. when $p = 3, 5, 11 \ldots$, the equation (8) cannot be satisfied if x, y and z are all prime to p.

The three congruences (10) can be replaced by

$$u^{2hp} \equiv (u^p + 1)^{2h} \equiv 1 \pmod{q},$$

or putting
$$u^p = w,$$
$$\left.\begin{aligned} w^{2h} &\equiv 1 \pmod{q} \\ (w+1)^{2h} &\equiv 1 \pmod{q} \end{aligned}\right\}.$$

By a detailed study of these last two congruences for special values of h, Dickson* showed in 1908 that the equation
$$x^p + y^p + z^p = 0$$
had no solutions in which the unknowns were all prime to p if $p < 7000$. Maillet had previously in 1897 proved the truth of this for $100 < p < 223$, while Mirimanoff in 1904 had raised the upper limit to 256†.

* *Quart. Journ. Math.* Vol. XL. 1908.

† More details of some of the results in this booklet are given in Bachmann's book *Das Fermatproblem* published in 1919.

ISBN: 0-8218-2674-3

9 780821 826744

CHEL/108.H

ISBN 0-8218-2674-3
9 780821 826744
CHEL/108.H